

A united Europe in 1992?

The European Commission plan to make Europe into an effective common market by 1992 is laudable, but seems like putting the cart before the horse.

THE plan that the European Communities should be made into a common market by the beginning of 1992 seems a little late. Has it not been the Common Market in the public mind for most of the past thirty years? The truth is that the grand designs of the Treaty of Rome for the economic integration of Western Europe have been either deliberately or tacitly postponed. The notion that people would be entirely free to move between jobs within the communities does not in practice apply to most professional people, whose qualifications may be recognized only in the countries in which they were trained or (irksome especially for academics) because many teaching and research jobs are counted as part of the public service and thus exempt from the requirement that employment should be undiscriminating.

Even on the strictly economic front, free movement is as often a promise as a reality. Many community governments (France, for example) retain restrictions on the export of capital, or are xenophobic in the welcome they offer to foreign attempts to purchase domestic assets (witness the French attempt to block the purchase of the financial newspaper *Les Echos* by the *Financial Times*). Other economic rationalizations are inhibited by continuing national and idiosyncratic regulation of particular industries (insurance and banking, for example) or by chauvinism in regard of "our shipbuilding/automobile/electronics industry". No wonder that Europe is full of doubters that these restraints will suddenly be swept away.

The European Commission's impatience with the continuation of these impediments to a truly common market is understandable, but the 1992 timetable it has set for itself and its member governments may be unrealistic. One proposal, for example, is that European governments should harmonize the rates of value-added tax applied to particular kinds of goods, which will lead to a dozen bitter domestic rows in each of a dozen countries about the propriety of taxing books, or wine. It is also proposed that border inspections of people's travel will no longer be necessary, as they already are within the United States, but that overlooks the continuing insistence of European governments that border inspections have a police and public health function. Community-wide institutions for meeting national governments' legitimate concerns would melt these objections.

Indeed, there is a sense in which the drive towards unification now under way is an attempt to achieve a desirable end by legislating for the external forms of unification, not the reality. This is the sense in which it makes no sense to talk of the European Communities acting as a single economic entity until they have a common defence policy, or even a foreign policy that they share. Even more strictly relevant to the economic issues now being tackled is the question of publicly sponsored research. Is it conceivable that the common market of 1992 will be one in which all economic entities compete on an equal footing if national governments are still free to pursue chauvinistic policies on research, fostering the health of particular industries by supporting research in strategic fields? Obviously not, but it is also unthinkable that governments will wind down their present arrangements by 1992.

Research deserves more attention than it has been given either by the commission or the member governments. Academic research, of course, would be exempt from restraint, even when governments exhort researchers to pay heed to the needs of national industries. But even here there will in the long run be complaints if governments take to complaining that they are doing more than their fair share of what will seem, after 1992, to be a common enterprise. Ultimately, there will have to be a common mechanism for supporting even basic work, preferably one that evolves from the present pattern of national research councils freed from restrictions (such as those requiring that grants be made only to their own nationals) supplemented by an enlarged community mechanism. The communities' existing CODEST operation, likely to be given more money now that the commission's budget has been settled, is still over-restricted by the requirement that its projects involve people from at least two countries. 1992 is not so far away that these questions can be left much longer.

The more serious problem concerns the direct support of industrial research, to which most European governments remain committed. The logic of the period after 1992 implies that support for research of this kind amounts to an unfair subsidy of industry, but governments are not about to kick the habit. Nor is there much conviction that the European solutions so far attempted, mostly by the commission, have been successful. Schemes for fostering collaborative development in fields such as information technology have mostly been successful at spending the money they have been allocated. There is much to be said for a self-denying ordinance, in which European governments agree that it is not their place to sponsor industrial research—but that will not happen by 1992.

The lesson to be drawn from this catalogue of contradictions is not that 1992 should be postponed again, but that European governments should recognize that the programme mapped out will create none of the common institutions needed. The more imaginative of them will begin now telling their fellow member-governments what needs to be done before the ambitions for 1992 can become an enduring reality. □

Squaring the circle

Dissension in the British government about economic policy has wider implications.

THE curious semi-public row that has broken out between the British Prime Minister, Mrs Margaret Thatcher, and her Chancellor of the Exchequer, Mr Nigel Lawson, should be instructive for others than the British. That the row centres on the apparently arcane matter of how a government should manage the external value of its currency does not diminish the importance of the disagreement or of the consequences there will be when it is eventually decided. That the British should regard the row as a sign that even an apparently invincible prime minister can encounter serious opposition over essentially political issues is natural, but unimportant.



In its simplest terms, the row is about the relative value of the British currency, sterling, and the West German deutschmark, now the most solid of West European currencies and the lynchpin of the European Monetary System (EMS) by means of which most European Community states (but not Britain) maintain the values of each others' currencies within narrow bands in a kind of regional simulation of the old global system of managed currencies set up at Bretton Woods and abandoned only in 1972. Briefly, Lawson wants to keep the value of sterling more or less where it has been since the collapse of the dollar last November, roughly DM3 to the pound sterling, and has been urging the Bank of England to sell sterling and buy deutschmarks to keep the rate constant. His motives are that he does not wish to see the competitiveness of British industry undermined in Western Europe and that he seems to recognize that Britain must sooner or later join EMS (no later than 1992 if moulding the European Community into a common market is not to be a nonsense).

Thatcher, on the other hand, is more concerned about the inflation rate, which has been marvellously reduced under her administration, but which, at just under 4 per cent a year, is still greater than in West Germany. But she appears also to be offended at the waste of public money involved in buying other people's currencies so as to depress the value of your own when, as she put it the other day, everybody knows that "you cannot buck the market". That proposition is well borne out by the experience of the years since 1972, but as stated overlooks the simple truth that one of the principal immediate causes of the recent strength of sterling is that British interest rates (also managed by the Bank of England, on the instruction of Lawson's Treasury) are higher than elsewhere, in West Germany in particular. The economical way of keeping the pound at DM3 would be to reduce interest rates, but that would allow inflation to increase.

Unsurprisingly, neither Lawson nor the governor of the Bank of England openly acknowledged that there has been a row behind the scenes when they gave evidence to the House of Commons Treasury and Civil Service Committee last week, which has not prevented observers parsing all their sentences in the hope of telling what the portents are. (Most probably, the row has now been patched up.) But it is a great surprise that two able and powerful people should find themselves at odds over such a simple matter in the management of a small industrialized state's economy.

The truth is that the prime minister and Lawson have declared two distinct and different objectives, but have more than two adjustable parameters at their disposal. Interest rates and the managed value of the currency are two of them. Another is the size of the government's budget deficit, which tortuously determines the supply of money in the economy, and which Lawson dealt with proudly in last month's budget by declaring that, for the future, the norm will be a balanced budget although, in the year ahead, the government will actually raise \$2,000 million more in taxes than it will spend. At the same time, Lawson brought about the biggest simplification of the British tax system ever at a cost of roughly \$4,000 million, most of that quarried from the extra taxes that would be paid on increased salaries in an increasingly prosperous community. But both the prime minister and Lawson know that spending less on tax reductions would have allowed interest rates to be reduced, thus helping the sterling to keep a constant value relative to the deutschmark without running the risk of inflation. There were, of course, political considerations why they agreed that the adjustability of this adjustable parameter should be mortgaged, but that is another story.

So, too, is the more worrying feature of the British economy represented by the deterioration of the balance of external trade and, worse, the secular decline of the savings ratio (the proportion of all after-tax income not immediately consumed) which has fallen from 14 per cent to just under 6 per cent since

1980—almost as bad as the ratio below 4 per cent in the United States. These two worrying trends are linked—less saving means more spending, much of it on imports. The lesson for the British and other improvident economies such as the United States is that governments should be more diligent in providing people with an incentive to save a more substantial fraction of their incomes. That is how Thatcher and Lawson, or their successors, must hope to patch up their differences. Unfortunately, considerations of that kind are also, for the time being, on the back burner. □

Windows copyright?

An impending copyright suit in the United States raises testing questions.

APPLE Corporation's suit against Microsoft (the software company) and Hewlett-Packard for alleged infringement of copyright on computer software raises interesting questions that, when decided, are bound to have far-reaching consequences. Apple's revival of fortunes with the introduction of the personal computer called Macintosh seems to have depended to a large extent on software innovations, not least a scheme for partitioning the display screen into separate patches called "windows", simultaneously usable for the display of information deriving from different programs. Unsurprisingly, other manufacturers of personal computers have taken up the idea (and even IBM is believed to be about to follow suit). But Apple is hoping to put a stop to that by suing its imitators for having breached its copyright. Apple has not made public the details of its claims, although it has talked about its right to protect the character of its method of presenting information on computer video screens and has said that its rivals' computer presentations "look too much like" its own.

The interest of this issue will turn on the degree to which Apple's suit may be judged a wish to protect by copyright law what is simply an idea, however useful and successful. Apple's sense of outrage at what its imitators are about might have been anticipated by the person who first had the idea of letting light into closed buildings by means of apertures in the walls as well as by the one who, later, thought it would make even better sense to let the light in and keep the cold out by covering the apertures with transparent material of some kind. Neither of these innovations would have been patentable under modern patent laws, although specific ways of manufacturing windows are patentable, and specific designs are subject to copyright protection. This is simply a restatement of the doctrine that ideas as such are not patentable.

That is how it should be. The entire system for protecting innovations would be unworkable if people were able to protect their ambitions to accomplish goals still unattainable. Since Apple is presumably legally well advised, it must therefore be assumed that it is not seeking to protect the notion that there should be separate windows on a computer screen, but merely the particular appearance of its own windows as distinct from those of other people, or even the particular methods by which its software has been designed to produce windows of different shape and appearance. If the software is the crucial issue, recent court decisions will give Apple a good case. If, on the other hand, it is asking that other computer manufacturers should be prevented from imitating the appearance of its own windows, it will have a harder case. It would then be as if *Nature*, having discovered that some other science journal had adopted its own design and layout, were to rush to the courts for redress. It would probably be wiser to take comfort in the maxim that imitation is the sincerest form of flattery—but to be prepared to sue only those who seek to pass off their imitations as the products they wish to emulate. That is what the law now requires. That is also what it should be. □

Controversial AIDS drug for US market

Berkeley, California

ETHIGEN Corporation of Los Angeles will market AL-721, a controversial lipid mixture reputed to have anti-AIDS action, as an over-the-counter nutritional supplement in US pharmacies and health food stores, beginning in April. The price will be \$250-\$300 for a month's supply.

Clinical benefits of AL-721, which is made from egg-yolks, have yet to be shown in controlled trials, but reports of its anti-viral properties resulted in a clamour from AIDS patients to make it available. Ethigen vice president Robert Weingarten said the company decided to begin sales now rather than wait for results of controlled trials in AIDS patients, due to increasing demand and a proliferation of 'copycat' compounds.

AL-721 was developed in the early 1980s by Meir Shinitzky at the Weizmann Institute in Israel, as a possible treatment for alcohol withdrawal and age-associated memory loss. It has not been proved effective for either condition, but a limited human trial suggested that it may improve immune function, spurring Ethigen, originally called Praxis Pharmaceuticals, to obtain the licence to produce it. Ethigen angered Weizmann officials by using the institute's name in promoting the unproved treatment.

AL-721 is Ethigen's first product, and the company's stock has ridden a roller coaster of publicity over the compound. Priced at \$1.25 per share when Praxis went public in January 1985, it jumped to \$4 per share that November, after publication in the *New England Journal of Medicine* of results from Robert Gallo's laboratory at the National Institutes of Health suggesting that the compound inhibited viral replication. In early 1987, an 8-patient, uncontrolled clinical trial in New York showed what Weingarten called "promising but inconclusive" reduction of virus in blood of some HIV (human immunodeficiency virus) patients with lymphadenopathy, boosting the company's stock to an all-time high of \$11 per share. Currently priced at about \$1.50 per share, the stock showed only a "muted response" to the recent marketing decision, said Weingarten.

AL-721 has been shown to scavenge excess cholesterol from cell membranes and from the AIDS virus envelope. Company scientists speculate that such a change could alter the configuration of surface proteins, preventing the virus from binding to cells. The company says it has found that the imitation compounds

New US-Japan science agreement reached

■ Cooperation to be expanded

■ Compromise on security issues

Washington

US AND Japanese negotiators at last concurred "in substance" last Thursday on the terms of a new science and technology cooperation agreement. It took eight rounds of talks, held alternately in Tokyo and Washington, to reach this point and neither side could manage a smile.

The agreement will be officially signed by US President Ronald Reagan and Japanese Prime Minister Noboru Takeshita at a time yet to be decided, but perhaps at the June economic summit in Toronto. Before then, legal details of the agreement will be worked out. The present bilateral science and technology agreement, which would have expired last Thursday, will be extended by three months to allow the finishing touches to be put to the new agreement.

Under the current agreement, 48 cooperative projects have been carried out in the fields of space, health, energy and environmental research. Included in the diverse areas studied are joint projects on Halley's comet, X-ray astronomy, typhoons, alcoholism and insect pest management.

The new agreement will not significantly affect projects already under way. But it will set much clearer terms for an expected major expansion of joint research by the two nations and provide a new management structure to make expansion possible. It is hoped that the number of US scientists working in Japan will rise as cooperation increases.

Although details of the new agreement remain secret, a major compromise ap-

pears to have been made on research that may affect national security. The Japanese government carries out only a little defence research, most of it at a Technical Research and Development Institute which employs just 1,200 scientists. Elsewhere in Japan, even in government laboratories, there is a strong tradition of open access to research results which Japanese researchers are anxious to maintain.

Early demands from US negotiators would have required a change in Japanese law to increase government control over joint research deemed by the United States to affect national security. Research related to the Strategic Defense Initiative would have been an obvious candidate for increased security.

The final agreement does contain a security agreement to prevent militarily-useful information from reaching other countries. But it does not require a change in Japanese law to enforce and is apparently worded in a way unlikely to outrage Japan's opposition parties.

Debate over how intellectual property rights from joint research are to be apportioned also appears to have ended in compromise. Instead of the new set of general rules that the US side is thought to have been seeking, rights will be given on a 'case-by-case' basis, according to the contributions made by each side. Although that is an advance from the current agreement, which provides no specific guidelines, it falls far short of a US view that all joint research conducted within the United States would be US property.

Alun Anderson

on the market lack the proper lipid ratio to produce the membrane effects seen with AL-721.

Despite doubts on the part of many AIDS researchers about the plausibility of the proposed action of AL-721, the National Institute of Allergy and Infectious Diseases (NIAID) has begun toxicity trials, due as much to pressure from AIDS sufferers as to any belief in its promise as a treatment. "It's not an obvious candidate to be a major breakthrough", said Dale Spriggs, of the AIDS treatment branch of NIAID.

Ethigen says it will comply with US Food and Drug Administration (FDA) guidelines, which allow a product to be marketed as a nutritional supplement only

if it is produced from substances recognized as safe, and the producer makes no claims of therapeutic effects. The FDA is currently investigating an Indiana company for marketing a compound similar to AL-721 as an AIDS treatment. Because of AL-721's widespread reputation, Ethigen is seeking advice from regulatory lawyers on whether it should make a specific disclaimer for AIDS.

AL-721 is currently available by prescription in Britain for patients with AIDS or AIDS-related conditions. It is classified as a "special" pharmaceutical, said Weingarten, which means it has not completed testing but has a high safety profile and some indication of possible efficacy.

Marcia Barinaga

Australia's researchers seek a break from reform

Canberra

THE pace of reform is causing problems in Australia's higher education system. Since the Minister for Employment, Education and Training, John Dawkins, issued a policy paper last December, institutions of higher education find they must concentrate scarce research funds in areas where they possess special expertise, and where they can serve the nation's economic needs.

The old system of distributing funds according to an institution's student enrolment and educational status has gone and, along with it, the Commonwealth Tertiary Education Commission. A new body, the National Board of Employment, Education and Training, directly answerable to Dawkins, is now responsible for giving out funds, through its sub-council, the Australian Research Council (ARC). An increasing proportion of those research funds will be given out through competitive bidding based on an institution's "educational profile". The profile is intended to set out the institutions' strengths and explain how they can best serve national priorities. A problem is that the government has so far identified only marine science as an area of priority.

Although ARC was originally conceived of as a statutory body within the Department of Science, it will now be instrumental in carrying out Dawkins' reforms. Its newly appointed interim chairman (and, so far, sole member) is Professor Don Aitkin, a political scientist at the Australian National University. He says that to expect the government to find the extra A\$50 million which in the Australian Science and Technology Council (ASTEC)'s view the ARC needs to be effective is unrealistic as "Australia slides into debt at \$1,000 million per month".

But another of Dawkins' reforms will provide him with an opportunity to find A\$30 million a year for the next three years. The current difference in level of funding between universities and colleges of advanced education is to disappear, and the extra funds that are now provided for universities are to be put up for competitive bidding.

With himself as the only member of ARC, Aitkin is concerned that there will be delays in getting the reforms off the ground unless Dawkins appoints the rest of the council within the next few weeks. The budget is due in July and the council needs to provide guidance to institutions so grant applications can be finalized before then.

Another important question still unan-

swered is who is to set Australia's research priorities. According to Aitkin, ARC, as the grant-giving body, will be the final arbiter.

Another candidate for the job is ASTEC which was established to advise the prime minister directly on matters of science, technology, and their effects on society. ASTEC is going through some changes. It has a new chairman, Professor Ray Martin, a research chemist who was vice-chancellor of Monash University in Melbourne until last year. And for the first time since its establishment, ASTEC is not working on any specific investigation for the prime minister. According to the secretary of ASTEC, Dr Greg Tegart, "Ray Martin has a clean slate". Martin wants ASTEC to play a role in setting Australia's research priorities.

Those priorities are certain to centre on aiding Australia's manufacturing industries. Department of Trade figures analysed by Tegart show that Australia lags far behind countries like Sweden and Norway in its volume of exports. With only sluggish growth expected in mineral and agricultural exports, Australia must turn to manufacturing.

Dr David Widdup, executive director of the Federation of Australian Scientific and Technological Societies (FASTS), doubts the ability of government to predict which areas of science and technology would be Australia's best bets and would prefer industry set priorities. He proposes that a one per cent levy be made on sales of manufactured goods and industry decide what research the money should be spent on. He believes the system has already worked well for many years in the coal, wool and wheat industries.

The government hopes that Australia's largest research body, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), will come to the nation's rescue as it has in the past, by acting as an engine to drive the growth in manufacturing that Australia needs.

CSIRO has just been through a major restructuring that resulted from a major report on it by ASTEC. According to CSIRO's chief executive officer, Dr Keith Boardman, restructuring was necessary to restore the balance between basic research and applied research. Basic research had been too much favoured.

But scrapping of CSIRO divisions and a shortage of funds has left morale low through much of the organization. CSIRO divisions are now expected to find 30 per cent of their budget directly

Living off the Earth

Washington

ALTHOUGH there are plenty of practical problems plaguing the US space programme, at least one US congressman believes that spiritual revitalization is also needed. To that end, Representative George Brown (Democrat, California) has introduced legislation that would establish human settlement of the space frontier as a long-range national objective, a goal he hopes will rekindle the national enthusiasm for space that helped to propel the Apollo lunar programme.

Extending human activities beyond Earth orbit has already received White House blessing in the new space policy signed by President Reagan in January (see *Nature* 331,550;1988). But most of the discussion about a national space policy has focused on shorter-term questions, such as whether to build the proposed permanently manned space station, or how best to reintroduce expendable launch vehicles to the US fleet.

Brown says the lack of long-range planning has contributed to the malaise that now afflicts the US space programme. The National Aeronautics and Space Administration (NASA) has begun to give some thought to returning to the Moon, or perhaps planning a manned mission that would go directly to Mars. His legislation would move such activities out of the back corridors of NASA headquarters and onto centre stage.

Some serious thought is being given to permanent communities off the Earth. Space Biospheres Ventures, a private company in Oracle, Arizona, is developing just such habitats. In January 1990, 8 people will enter Biosphere 2, an enclosed structure covering more than two acres with a volume of some 5 million cubic feet. Inside there are a variety of biomes; a wilderness area, a tropical rainforest, a savannah, an ocean, salt and freshwater marshes and a desert.

from industry. According to Dr Graeme Caughley of the Division of Wildlife and Ecology, the division's budget has fallen 5-10 per cent in real terms each year for several years. Operating expenses cannot be cut further and staff layoffs will be the only possibility if further cuts are made. By its nature, much of the division's work has no appeal to industry, unless perhaps to the timber industry, whose money, Caughley professes, his colleagues would hesitate to accept because of a possible conflict of interests.

Boardman summed up CSIRO's position with the statement, "What we need now is 3-5 years of stability".

Charles Morgan

The mix of ecosystems should make it possible to sustain life for two years in a totally enclosed environment. Such a project could be the prototype for artificial ecosystems on other planets.

Others have speculated on the macroeconomics of human activities off the surface of the Earth. The Space Studies Institute has considered using raw materials mined on the Moon for constructing Earth-orbiting satellites. One project of particular interest is constructing power satellites capable of converting sufficient solar energy to provide an economically viable supply of energy to Earth.

Eric Jones of Los Alamos National Laboratory has also studied how human settlements might make use of locally available resources on the Moon. Jones draws the analogy of early settlement of the American West and Australia. For these pioneers, their local environment was a rich source of raw materials, and so too could be the Moon, with its favourable gravity making transportation to Earth relatively easy.

Were it to stand alone, Brown's legislation would have little chance of passage. But Brown hopes to amend it to the NASA authorization bill. If he can infect other congressmen with his enthusiasm for settling space, NASA will have to start recruiting for settlers instead of astronauts.

Joseph Palca

Conflict continues between Armenians and Azerbaijanis

London

THE current conflicts between Soviet Armenians and Azerbaijanis over the autonomous region of Nagorno-Karabakh has highlighted one of the major faults of pre-*perestroika* planning; the ease with which mistakes and failures could be concealed by gross-figure 'achievements'.

Although the petition by the Armenian population of Nagorno-Karabakh to be administered from Erevan rather than Baku is also based on considerations of cultural identity, economic factors have a particular relevance. For Lenin's original decision in December 1917 was that, although not contiguous with Armenia, Nagorno-Karabakh was "irrevocably" part of Armenia. In 1923, however, the region was assigned to Azerbaijan, on the grounds that this would ensure its more rapid economic and social development. So, on 11 February 1988, when the petition was launched at a rally in Stepanakert, asking for the region to be returned to Armenia, the signatories were in effect challenging the legitimacy of the 1923 decision. They maintained that Nagorno-Karabakh was economi-

cally disadvantaged in relation to the rest of Azerbaijan.

The Azerbaijan authorities organized a high-level 'round table' to refute these claims. Their deliberations were eventually reproduced in the party daily *Bakinskii Rabochii* and indicated that the population of Nagorno-Karabakh was significantly better off in social amenities, not only than the rest of Azerbaijan, but also often Armenia, and even than the Soviet average.

Crude figures can, however, be deceptive. The table presented by *Bakinskii Rabochii* gives Nagorno-Karabakh the advantage over Armenia according to eight of the nine chosen indices, and over Azerbaijan as a whole according to seven indices. But a reading of the figures suggests that this is not the whole story. The excess of middle-level medical staff (Nagorno-Karabakh 122.7 per 10,000 population, Azerbaijan 93.5, Armenia 93.5) coming immediately after the figures for doctors (Nagorno-Karabakh 29.1, Azerbaijan 38.4, Armenia 86.6 per 10,000 inhabitants) suggests that extra nurses and paramedics may have been drafted in to make up for the shortage of doctors. And the figures for clubs, cinemas and libraries, more than twice the average for either of the two contending republics, may be due to duplication of language facilities.

By the time these figures appeared in print, however, experts in Moscow were speaking of "serious planning errors" regarding Nagorno-Karabakh. The *Bakinskii Rabochii* data were "precise", it was now explained, but did not give the full picture. Higher education in the region, for example, consists of a single teachers' training institute—which, in spite of the almost 80 per cent Armenian population, offers no courses in Armenian history or geography. (Reserved places at Armenian universities have now been promised to students from Nagorno-Karabakh.) Another problem in Nagorno-Karabakh, the lack of jobs for school leavers and the consequent drift of young people from the region, seems associated with the failure of the Baku planners to ensure adequate roads, water and gas supplies to Nagorno-Karabakh, without which local agriculture and industry are unable to expand. And the lack of understanding of the cultural-ethnic tensions which have developed between the Armenians and Azerbaijanis in the region are yet another symptom of the Soviet Union's acute lack of people trained in demography and associated disciplines (see *Nature* 331, 555; 1988).

Vera Rich

Congressional roasting for NIH critics

Washington

THE US Congress's love affair with the National Institutes of Health (NIH) reveals itself in the strangest ways. Last week, NIH director James Wyngaarden found himself under fire from Representative John Dingell, the chairman of a subcommittee that oversees NIH activities, for having yielded to arm-twisting when he downgraded the status of three of NIH's research administrators from "outstanding" to "excellent" status.

The person doing the arm-twisting, Thomas Burke, chief of staff of the Department of Health and Human Services (HHS), was roasted by Dingell at the same congressional hearing. Burke protested that he had simply been trying to reduce the number of people with the "outstanding" rating; some HHS departments had come to contain nobody who was not "outstanding"—a strange contradiction in terms. But Dingell clearly felt that such action would damage morale at NIH.

Dingell is among the congressional supporters of NIH who are deeply concerned, as he put it, that "top-level researchers and managers no longer find the NIH an attractive place to work". While many at the Bethesda campus might rate NIH's charms higher, there is no doubt that the difference in pay between NIH and the universities is now perceived by Congress

and Wyngaarden as a major threat to NIH's continued preeminence.

Wyngaarden produced figures showing that US medical school base salaries for department chairmen range from \$109,000 in basic science to \$179,000 in clinical science. Their counterparts at NIH receive \$82,000, and there are strict federal government limits on the amount of bonuses that can be paid. The biggest worry, according to Wyngaarden, is not that they will simply lose established researchers to the universities, but that they will fail to attract the brilliant young researchers needed to maintain NIH's reputation over the next 10–20 years.

NIH have suggested legislation that would create a 'senior biomedical research service', which would free federal biomedical research organization researchers from civil service pay limitations. The Office of Management and Budget has suggested going one step further and "privatizing" NIH Bethesda campus, turning it into a university or research foundation outside the government. Neither view is likely to prevail in the near future, according to HHS officials at the hearing. Instead, all parties are likely to wait for the results of a report on NIH's prospects commissioned from the Institute of Medicine and due to be completed in November. Alun Anderson

European Community plans to sequence yeast genome

Munich

WHILE the United States National Institutes of Health are attracting the world's attention with their grand plans to map and sequence the human genome, the European Community (EC) has quietly begun a smaller but still significant project: to sequence the genome of yeast.

A two-year pilot project to sequence one yeast chromosome was passed in mid-February in Brussels by the advisory committee of the EC's Biotechnology Action Programme. The project will provide 2 million ECU (1 ECU = \$1.24 at current rates) to between 20 and 40 laboratories. Formal approval from the EC is expected soon.

The project could be expanded in 1991 under the BRIDGE programme to sequence the entire yeast genome, in conjunction with Japan and the United States. Project leader Andre Goffeau said the EC might make as much as 10-20 million ECU available for the sequencing of yeast and other 'industrial microbes' under the auspices of BRIDGE.

Fifty yeast laboratories from almost every EC country—but especially from France, West Germany and Belgium—have expressed an interest in the pilot project. Each laboratory would receive 10-15 kilobases (kb) per year to sequence using methods of their own choosing. In addition, each laboratory would check

the results of one other laboratory.

The chromosome to be sequenced, chromosome 3, was chosen because of its small size as well as because of access to an organized library of overlapping clones thanks to Maynard Olson of Washington University in St Louis and others in the United States. Chromosome 3 contains about 360 kb of DNA; the entire yeast genome contains about 15,000 kb. By contrast, the human genome consists of 3 million kb.

A centre to coordinate the sequence data is essential, says Goffeau. A likely candidate is the Martinsried Institute for Protein Sequence data near Munich, West Germany.

Goffeau is optimistic that the yeast sequencing project will attract widespread industrial support, which may be essential for the project's well-being. A preliminary study revealed broad interest from breweries, pharmaceutical companies and biotechnology firms in five or six EC countries.

Goffeau sees the yeast project as a stepping stone to the sequencing of the human genome, which will take much longer because of the lack of an organized library. By the time systematic sequencing of the human genome begins, "we will have learned a lot from yeast about interpreting sequence data."

Steven Dickman

Harvard tries teaching from afar

Boston

A HARVARD professor last week taught a method for solving a linear algebra problem not only to her students in the United States, but at the same time to a faculty panel in Beijing, China via an interactive computer and telephone connection. The so-called simultaneous classroom was a preview of a computer communication system to be used for a five-week course between Harvard and Beijing Normal University on expert systems that will begin next month.

The system employs a communication link that allows voice and data messages to be sent over a single telephone line simultaneously. The teacher, the US students and the Chinese delegation each use workstations with IBM-compatible computers, speakerphones for voice communication, and electronic tablets so equations can be written by hand.

The preliminary two-hour segment was, for the most part, a demonstration of the US style of teaching mathematics, where interaction between students and faculty is more frequent than in the Chinese system. But the first session included a fair amount of ceremony as well. While still video images depicting each campus flickered on the computer screens, the two university presidents introduced their respective institutions with short, if lofty, presentations.

The classes on expert systems will use the same system to link Harvard computer scientist Henry Leitner with approximately twenty Chinese graduate students in computer science, using educational software developed by a firm called Cambridge Teleteaching Group.

The Harvard programme comes hard on the heels of an arrangement China has made with Boston University Medical School to use a similar system to teach a five-week series on medical topics. Both of these programmes were initiated by Chinese officials who contacted the US State Department hoping to strengthen educational and informational ties between the two countries. The teleteaching programmes received a small grant from the US State Department, with the telephone time donated by AT&T.

A teleteaching experiment between the United States and the Soviet Union was launched last month at Tufts University using an interactive video link. This course on nuclear history pairs Professor Martin Sherwin and class with a Soviet group headed by Yevgeny Veli-kov, vice president of the Soviet Academy of Sciences.

Seth Shulman

UK wind power project

London

BRITAIN'S Department of Energy and Central Electricity Generating Board have announced a joint \$30 million pilot programme to produce electricity from three experimental 'wind parks' and a single offshore wind turbine. Each park, whose sites have yet to be chosen, will have 25 wind turbines, generating sufficient electricity for the needs of 15,000 people. According to Lord Marshall, the board's chairman, a typical wind park could consist of 30-m-high turbines of either vertical or horizontal axis design, with blades of about 30-m diameter. Each machine would have an output of 300-500 kW and a single park would occupy 3-4 square kilometres. The intention is that the wind parks should be built within the next four years, with a 25 MW output by 1993. The offshore turbine is expected to have a horizontal axis mounted on a steel tripod some 8 km off the East Anglian coast.

When the government announced its plans for the privatization of the electricity supply industry last month, a commit-

ment was given for the new distribution companies to place contracts for a minimum (as yet unquantified) proportion of non-fossil-derived electricity, mainly as a way to guarantee the continuing expansion of the nuclear power programme, but also to encourage a greater diversity of sources of renewable energy, such as wind power.

Launching the wind park programme on 23 March, energy minister Michael Spicer estimated that upon the successful development of the technology, wind energy could contribute up to 20 per cent of Britain's electricity. He conceded, however, that society's attitude to the prospect of large numbers of wind turbines spread over vast tracts of the countryside would determine the eventual feasibility of the scheme. It has been calculated that on the basis of present technology, it could take a hundred square miles of windmills to produce one per cent of the nation's electricity. "It is essential, therefore, that we consider very carefully the environmental implications of pursuing this technology", said the minister. Simon Hadlington

Scrutiny of defence spending reveals problems in UK labs

London

BRITAIN'S defence research establishments are suffering from serious shortages of resources and manpower, according to a government-commissioned study into the management of defence procurement. But because no useful method exists for measuring the shortages, no accurate analysis can be made and no meaningful recommendations for action can be suggested, the report of the inquiry says.

The investigation was launched following the cancellation by the Ministry of Defence of the General Electric Company's Nimrod early-warning radar system in December 1986, after some \$700 million had already been spent.

The government has agreed to make substantial changes to the way defence contracts are procured and managed after the report disclosed that the Ministry of Defence annually overshoots its procurement estimates by £3-4,000 million, on a budget of £9,000 million.

The report, which was originally confidential but declassified last week, condemns the ministry for its persistent failure to control expensive projects. A lack of clear definition of responsibilities and inadequate monitoring of the technical feasibility of equipment are cited as two reasons why present arrangements do not "provide a degree of control commensurate with the complexity of the task and the sums . . . being spent".

The report's main recommendations, which have been accepted, are that all major projects should be proceeded with step by step, so that the technology can be fully assessed and full development undertaken only when clear performance goals and acceptance criteria can be established and made binding on the contractor, and that each major project should have a single project manager with control of all relevant resources, including specialist technical support.

The report makes it clear that confusion exists over the precise role of the ministry's research establishments; "although all-embracing descriptions of their activities exist, we have been unable to find a clear statement of their role which identifies functions according to a set of priorities".

The report suggests that the establishments view their role primarily as medium-and long-term and systems research, with the ability to do technical evaluations and contribute to managers' decision-making only as a spin-off. A clear definition of the establishments' various roles is needed, the report says, with a distinction between work done to sup-

port the ministry's equipment procurement and work done to advance the technology base.

Within the research establishments, the investigators found widespread dissatisfaction with the levels of staffing and resources. Overall vacancy figures for the non-nuclear research establishments showed that around 10 per cent of posts were unfilled at the higher scientific officer level. The vacancy rate in electronics and computer science was running at 15 per cent. The investigators disclose that in one research establishment department specializing in software, staff numbers had declined from 60 to 45 over the past 2-3 years. "And we came across at least one case where a civil servant had

resigned on Friday to join a software consultancy and been re-employed (at higher cost) on the following Monday to do the same work". Of 14 project managers interviewed, nine felt they were getting insufficient support from research establishments and three of the remaining five were seriously concerned about future levels of support.

The report tells how one establishment director followed up his list of statistics with a letter emphasizing "in the strongest possible terms" how the statistics failed to show the "extreme problems" his establishment faced.

The authors of the report accept that "there is indeed a serious problem in staffing in key areas" but that "the nature of the problem remains far from clear". What are needed are management systems that can collect information that can be used as a basis for action.

Simon Hadlington

Keeping secrets is a bad idea in the superconductor race

Munich

KEEPING a secret has become tougher than ever—especially in the field of high-temperature superconductivity. West German pharmaceutical heavyweight Hoechst AG learned this the hard way two months ago, when a Japanese group announced the discovery of superconductivity at 85 and 110 K in a ceramic material very similar to one patented by Hoechst two months earlier.

The compounds represent a new departure from the first superconducting materials, discovered at the IBM labs in Rüschlikon, which were shown to conduct current without resistance at these temperatures. The compounds substitute bismuth for yttrium and might therefore be much less expensive to produce on a large scale. (Although bismuth is less abundant in the Earth's crust, it is a by-product of copper and zinc smelting.) None of the superconducting compounds has so far become commercially useful (see *Nature* 332, 305; 1988).

Hoechst missed out on the flurry of publicity that greeted the Japanese group, led by Hiroshi Maeda at the National Research Institute for Metals in Tsukuba, when it announced its discovery on 22 January 1988. University of Houston researcher Paul Chu also attracted attention when he announced an analogous finding three days later.

The West German group, which included workers at the Max Planck Institute (MPI) for Solid State Research in Stuttgart and the Technische Hochschule at Aachen as well as at Hoechst in Frankfurt, had found superconductivity in a bismuth-containing compound as

early as April 1987, but the results could be reliably reproduced only in October. The crystal structure of the West German compound will appear in the journal *Angewandte Chemie* in April 1988.

In a field where companies such as IBM and AT&T Bell Laboratories have encouraged their researchers to publish early and often, Hoechst's attitude seems anachronistic to some of the researchers concerned, who favoured immediate publication.

Hoechst research director Heinz Harnisch admitted only a slight change of heart. "In general, we still think our economic gains from not publishing are worth a bit more than good publicity." But the experience with this superconductor showed that early publication "might be advantageous".

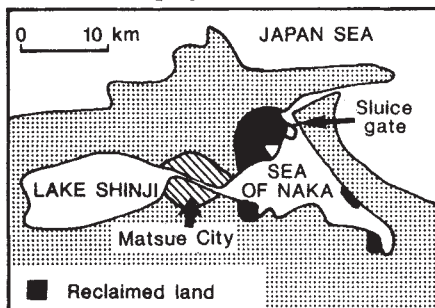
Hoechst will share whatever secrets it has left in superconductivity research with two other West German conglomerates. A collaboration among Hoechst, Daimler-Benz (the maker of Mercedes automobiles) and Siemens, as well as about a dozen West German universities and research institutes, was announced last week. Hoechst research director Heinz Harnisch said that cooperation is the "only way for West Germany to compete" in this field, especially against the Japanese. Hoechst will concentrate on developing superconducting materials while the other two firms will focus on applications. The development is sure to please the West German government, which has lately encouraged such collaboration. The government might even add some funding to the industrial programme.

Steven Dickman

Plans for giant freshwater pond opposed by Japanese

Tokyo

TWENTY years ago the Japanese Ministry of Agriculture, Forestry and Fisheries started a project to turn the inland



Sea of Naka and a brackish lake on the Japan Sea coast into a gigantic freshwater pond for watering rice paddy fields that were to be established on land reclaimed from Naka Sea. In the immediate postwar period when there was a food shortage in Japan, the plan made some sense, but two years after the project got under way a rice reduction policy came into effect to curb rice over-supply. Nevertheless, the ministry has forged ahead and sunk hundreds of millions of dollars into land reclamation and the building of a gigantic set of sluice gates at the mouth of Naka Sea. And next September the ministry plans to close the gates. But conservationists, fishermen and a majority of the local people are opposed to the plan because it will destroy a productive local fishery and will probably have devastating environmental effects on one of Japan's few remaining brackish lakes.

The plan to "desalinize" Naka Sea and Lake Shinji on the border of Shimane and Tottori Prefectures dates back to the 1920s when the channel linking the lake and Naka Sea was widened and deepened to allow passage of ferries. Seawater and marine life flooded into the already brackish lake and the local fishing industry boomed. But farmers complained that the saltier water was damaging their rice fields and they called for the mouth of Naka Sea to be dammed. However, it was not until the advent of a food shortage after the war that the ministry began to lay plans and the project finally got under way in 1968.

Over 2,500 hectares of land have been reclaimed at a cost of 72,000 million yen (\$576 million). And the ministry has won another ¥6,000 million in fiscal year 1988 to close the gate and develop some of the reclaimed land—apart from a trial attempt to grow a few cabbages, the land remains unused.

Shutting off the supply of seawater will destroy a local fishery for sea bass and

shimeji clams that earns more than ¥3,000 million (\$23 million) a year. There are also fears that thousands of wild duck and geese which overwinter around the lake and sea will be endangered because they feed on the tiny shimeji clams that abound in the brackish waters of Lake Shinji. And restriction of circulation and accumulation of nutrients may lead to suffocating algal blooms, such as occur in a former brackish lake north of Tokyo that was similarly shut off from the sea.

In February 10,000 people linked hands around Lake Shinji in protest and opinion polls indicate 70 per cent opposi-

tion to the ministry's plan to close the gate. During recent budget deliberations in the Diet, Prime Minister Noboru Takeshita who represents the Shimane area said that he "cannot ignore the reality that the local people's need has changed greatly since the start of the project". And the secretary general of the Shimane Prefecture association of the ruling Liberal Democratic Party has suggested that some of the reclaimed land could be converted into a golf course.

But, apart from a compromise proposal to leave one of the ten sluice gates open for a trial period, the ministry is determined to shut the gates. As one Shimane University professor put it, Japanese public works projects are run like the Japanese army; once the advance begins there can be no retreat.

David Swinbanks

Leningrad library loses irreplaceable publications

London

ALMOST 400,000 books and periodicals, including rare and irreplaceable works, were destroyed in a fire last month at the Leningrad library of the Soviet Academy of Sciences. At the time of the fire, the library had been without an operational fire-alarm system for over four years. It was due to be coupled up to a new automatic alarm network on 1 February 1988. Unfortunately, the date was postponed when an engineer fell ill. On the evening of Sunday, 14 February, fire broke out in the stacks.

The fire initially attracted little public attention. Immediately after the fire, the director of the library, presumably in a state of shock, estimated the financial loss at about 3,000 roubles (£3,000). On the basis of this, TASS announced that there had been partial damage to the newspaper holdings, but that "the country's treasure, the main archive of academic editions, rare manuscripts, exchange and specialized stocks and many other valuable items remained intact". Later investigation by L. Tereshkin, a reporter from *Leningradskaya Pravda*, revealed that there was also considerable damage to the book holdings, including the famous Baer collection, bequeathed to the academy by the zoologist Karl Ernst Baer (1792-1876). Later estimates put the loss at 300,000 roubles, but even this, library staff feel, is almost certainly too low.

Almost half the works lost are of foreign origin, so that inflation will have an enormous effect on replacement costs. Only very rough estimates (by metres of shelving) have been made so far of the books damaged by scorching or water, and which will require extensive repair

and salvage work. Academician Dmitrii Likhachev, the mediaeval historian who led the campaign against the diversion of the Siberian rivers, warned the annual general meeting of the Soviet Academy of Sciences last month that "the most urgent and resolute measures" will be needed if these books are to be saved. He estimated the priceless books lost as "several tens of thousands".

These works, however, Tereshkin says, have not been included in the estimated financial loss. Books and manuscripts which are "priceless", he says, are not even assigned a nominal price in the inventory, so that, he said, the library could have lost all its ancient manuscript collection and, according to the book-keeping now in force, have suffered no loss at all. The party and trade union organizations at the library have asked for a state commission to be appointed to produce an "objective" estimate of the losses. Leningrad's fire chief, Lieutenant-Colonel A.G. Ivanov, is also unhappy. Hardly any of the historic buildings and scientific collections within his precinct have proper fire alarm systems. The Institute of Archaeology, the Institute of Oriental Studies, the Academy of Fine Arts, the Cathedral of Saints Peter and Paul, the Zoological Museum, and the Museum of Ethnography, have no alarm system at all, while others have only outdated and non-automatic systems. In 1986, he told Tereshkin, there was a major fire, in which several lives were lost, in the Yusupov palace. There was no alarm system at the time of the fire, and none has yet been installed. How much more will be needed, he asked, to rouse the city authorities from their complacency? Vera Rich

The thesis that won't go away

SIR—Beverley Halstead (*Nature* 331, 497; 1988) appears to favour the abolition of the PhD thesis and its replacement with mainly published work. There are, however, several potential shortcomings of a method based solely or largely on publications.

A paper does not necessarily convey the amount of work, or more importantly, the scientific thought and method required to obtain the published results. Those working to establish a new research project or a novel method may demonstrate greater research prowess but produce less publishable material than subsequent workers who can build upon these foundations. A thesis should not contain every failed experiment but it should provide a forum for areas of work that are not publishable.

An equally serious objection is that a student with a long list of publications may obtain a PhD without having actually reached the required standards. This may sound paradoxical, but if the candidate is working in a large, well-established research group, there exists the possibility of the student being carried along by the group as a whole. In these days of papers with several authors, how is the examiner to assess the contribution made to the research by the candidate?

I believe that the PhD thesis in (essentially) its present form does still have a major role to play. But there should be a move away from the belief that thesis weight is directly proportional to scientific quality and students should be directed towards brevity and clarity in their writing.

Moreover, having just recovered from the trauma of completing a (relatively short but as yet unexamined) thesis, I would hate to think that all that effort was a waste of time.

RICHARD CANNELL

*Division of Biological and
Environmental Sciences,
Hatfield Polytechnic,
Hatfield, Herts AL10 9AB, UK*

SIR—It is generally recognized that a doctoral thesis is in want of a readership. The Swedish university system, however, offers definite advantages over the British. Indeed, it satisfies most of the points raised by Beverley Halstead.

In Sweden, a thesis is normally based on four or five articles published in major international refereed journals. These are bound as part of the thesis. The candidate also writes a review, which defines the problem, surveys the literature and discusses the work in context. Practical details and methods described in the papers are not reiterated. The thesis is given an ISBN number, and printed and distributed at the faculty's expense. Doctoral theses in

scientific subjects are rarely based on unpublished work.

My review was about 8,000 words, excluding references. This is average. There have, however, been shorter essays, and candidates are encouraged to keep to the point. Also, before presenting a thesis candidates have, among other things, to attend practical and lecture courses.

The candidate must physically nail the printed thesis to a specified wall or door of the university. This mediaeval ceremony, in a country not obsessed with tradition, conjures up Martin Luther.

The oral defence of a thesis is a public proceeding. There were probably fifty people at my disputation. The opponent (external examiner) summarized the thesis and proceeded with questions. The other examiners also joined in.

Up to about 25 years ago, a thesis had to be original work. The defence could be a very fiery affair — truly a disputation. There were three opponents. The quality of the doctorate was judged not only on the written thesis but also on its oral defence. Some of these theses, for example Robin Fåhræus's "The suspension stability of the blood", are scientific classics. The system was revised after the war because of changing research patterns.

Thus many of the reforms which Halstead advocates can be observed in the Swedish doctoral system, which is worth studying before the British system is revised. But Swedish doctoral students have no illusions that their circle of readership is any wider than in Britain.

LARS H. BREIMER

*Institute of Cancer Research,
Chester Beatty Laboratories,
London SW3 6JB, UK*

SIR—Beverley Halstead's arguments against PhD by thesis suffer from an horizon constrained by the university.

If 50 per cent of first-degree candidates failed there would be an outcry. But a loud and prominent part of it would be a burst of cheering from technical interviewers of candidates for employment in industry, who at present conclude that many graduates should have failed.

Second, most PhDs are not going into academic research and the publication of academic papers. Even of those who continue in research, most will go into industry where they write reports, not papers, and proposals for further action. By the standards of an effective report, academic papers are intolerably verbose, indefinite and indigestible. Papers are full of hedges and provisos, ifs and buts, and erudite circumlocutions; effective proposals are not. Moreover, the approved style for papers is intolerably dull. It would be very surprising if Halstead's

arguments could not be put more forcefully in a third of the words he has employed.

Writing a cogent argument would be useful training for most PhDs, writing a typical learned paper is not. And the inevitable result of Halstead's proposals would be more papers, in more specialist journals with even lower editorial standards. It takes a lot to make an external examiner fail a thesis; if papers were the criterion, editors and reviewers would be under the same moral pressure.

Do even the academics want ever more trivial and unreadable 'learned' papers?

SIMON ROMAN

*2 Upper Rosemary Hill
Kenilworth CV8 2PA, UK*

SIR—The discussion by Beverley Halstead prompts a related question about the function in Britain of the PhD studentship, especially because of the recent trend to employ research assistants with the specific guarantee of a PhD registration. This trend has invidious and undesirable consequences for the unity and morale of PhD students in a department and also for the training of young scientists for independent research.

There are now two classes of students: those who exist on a grant of about £3,500 a year, and those who live in relative comfort on approximately twice as much disposable income. This situation is divisive and illogical. Many of the best graduate students compete for studentships, only to find subsequently that less well qualified individuals have been offered the same opportunity to complete a PhD, but as paid research assistants.

This leads to resentment and outright disbelief in the system. The chief problem is that roles are confused, and not simply in the unfortunate financial discrepancy. The PhD student has traditionally been considered as one who, while contributing to the overall scientific achievement of his or her group and department, is obliged to pursue independent research at the highest level, free from the constraints of grant applications, teaching duties and, most importantly, who is able to pursue a scientific theme without fear of being reprimanded by short-sighted employers for irrelevance or lack of results.

Traditionally, research assistants have been considered as apprentices to this form of training. In practice, the research assistant PhD student is often subject to constraints that are at odds with the true concept of the PhD degree. His or her project may be geared solely towards the benefit of the group, possibly involving repetitive, incoherent and tenuously related pieces of research, wholly inappropriate for a PhD thesis.

The undoubted discrimination, both financially and functionally, between the PhD studentship and the research assistant-

ship calls into question the PhD degree as a unified system of training and examination. Those responsible for the trend must decide whether there is a role for the studentship, and if the answer is positive, they should endeavour to fight for a better grant for the PhD student, and restrict PhD research assistantships to exceptional cases. If the answer is paid employment of all PhD students, those responsible should remember that it is their responsibility to promote individuality, innovation and a sense of independence, qualities which have undoubtedly been part of the reason for their own elevated position. If the current situation continues, the institution of the PhD student can only be undermined and possibly lost.

A.J. GREENFIELD

81 Wellesley Court,
London W9 1R1, UK

SIR—With reference to Beverley Halstead's comments on PhD training, the thinking at South African universities is currently in a 'reform' phase. A common view among academics is that the thesis should still weigh in at a certain level (it must not be too light) but that it is desirable that at least some of the material should have appeared in print before submission of the thesis. A reason often given is that papers published in refereed journals represent a sort of antidote to over-critical examiners. What has recently stimulated supervisors to encourage MSc and PhD students to publish is, however, the fact that the annual government subsidy now received by each university is based partly on the publication output (in high-quality journals) of the institution concerned. Some of this money frequently finds its way into the supervisor's research grant.

MILES B. MARKUS

Department of Zoology
University of the Witwatersrand,
Johannesburg, South Africa

SIR—The commentary by Beverley Halstead emphasizes the importance of publishing parts of the thesis before the PhD is awarded, a view I fully support. However, the space permitted for a publication often varies between journals, the most contested journals generally requiring the author to be particularly brief and concise. As a consequence, a published paper may not contain the wider discussion of a subject as is expected from a thesis, and hence the two forms of written work should ideally complement each other. In this respect, I believe that the British PhD is more than a collection of original (published) papers and represents an important contribution to the scientific literature in its own right. Halstead writes: "During the course of my PhD I published a paper — it was far more exciting for me to do this than write a thesis". To this I

would like to reply: during the course of my PhD I published a paper — and as I was restricted to keeping the length of the paper to the absolute minimum I particularly enjoyed writing the corresponding chapter of my thesis. The paper was published, but the thesis gives a fuller account of the scientific findings and their relevance.

Is it the relative inaccessibility of the British thesis that needs a change? At present, only one mandatory copy is deposited in the departmental and in the main library of the university where the degree is taken (in my case at Reading in England). The printing and circulation of a large number of copies, traditionally the case in West Germany, would improve this situation.

G.H. SCHMIDT

Department of Cell Biology,
Fraunhofer-Institute, ITA,
Nikolai-Fuchs-Str. 1,
3000 Hannover 61, FRG

Turin Shroud

SIR—Following my letter of 11 June 1987 (*Nature* 327, 456; 1987), I am now able to provide an outline of the procedures that have been finally agreed for the radiocarbon dating of the Shroud of Turin. Of the seven original offers to undertake the dating of the Shroud, three have been accepted by Cardinal Ballestrero, Archbishop of Turin and Pontifical Custodian of the Shroud. The radiocarbon laboratories concerned are at the University of Arizona, the University of Oxford and the Federal Institute of Technology in Zurich, and each has now agreed to proceed with the project.

Each laboratory will be provided with a sample from the shroud, together with two known-age control samples, one of which will have been independently dated by conventional radiocarbon dating. The shroud samples will be taken from a single site on the main body of the shroud away from any patches or charred areas. In order to ensure that ample carbon for dating survives after pretreatment, the weight of each cloth sample (that is, shroud and controls) will be 40 mg. All the samples will be given to the laboratories as whole pieces of cloth without being unravelled or shredded. A blind test procedure will be adopted in that the three samples given to each laboratory will be labelled 1, 2 and 3 and the laboratories will not be told which sample comes from the shroud. Even if the samples were shredded, it would still be possible for a laboratory to distinguish the shroud sample from the others. It is therefore accepted that the blind test depends ultimately on the good faith of the laboratories.

The removal of the samples from the shroud will be undertaken under the supervision of a qualified textile expert. These samples will be weighed, wrapped

in aluminium foil and sealed in numbered stainless steel containers. The control samples will be similarly treated. All these operations will be watched over and certified by Cardinal Ballestrero in collaboration with myself. After they have been packaged, we will immediately hand over three samples (shroud and two controls) to representatives of each of the three radiocarbon dating laboratories who will be in Turin for this purpose. In addition, all stages will be fully documented by video film and photography.

On completion of their measurements, the laboratories will send their data for the three samples to both myself at the British Museum and to the Institute of Metrology "G. Colonnetti" in Turin for preliminary statistical analysis. The laboratories have agreed not to discuss their results with each other until after they have deposited their data for statistical analysis. A final discussion of the measurement data will be made at a subsequent meeting in Turin between representatives of these two institutions and representatives of the three laboratories at which the identity of the three samples will be revealed. The results as finalized at this meeting will form a basis for both a scientific paper and for communication to the public. The timetable for the operations has not yet been fully established but it is hoped that a radiocarbon date for the Shroud of Turin will be released by the end of 1988.

M. S. TITE

Research Laboratory,
British Museum,
London WC1B 3DG,
UK

AIDS and HIV

SIR—In Gregorio Weber's letter (*Nature* 330, 690; 1987) he says: "The prolonged absence of symptoms in a large majority of those that have contact with the [AIDS] virus is already indicative that in many — probably most — of them the disease will never develop."

In fact, the San Francisco AIDS Research Study (director Paul O'Malley) has published findings that after seven years, 36 per cent of those infected with HIV (human immunodeficiency virus) have developed AIDS and another 38 per cent have developed AIDS-related complex. These results were released at the Third International Conference on AIDS in Washington in June 1987.

It is true that early in the epidemic health officials thought (or hoped) that only 10 per cent of those infected would manifest this illness. Time has shown that this is not the case.

DONALD J. GORMAN

DAIR Foundation, Inc.,
2336 Market Street, Suite 33,
San Francisco,
California 94114, USA

Is the Earth alive or dead?

James Lovelock's Gaia hypothesis seems to have had a fair hearing at a meeting in California last month. It may matter less whether the theory is correct or otherwise than that it is stimulating.

San Diego

In few scientific conferences is the most talked-about session a debate, lasting into the night, on the philosophy of knowledge and the gradations between theories, hypotheses and mere ideas. These are subjects generally as interesting to scientists as the manufacture of concrete is to architects. Yet many of those attending the American Geophysical Union's San Diego meeting (7–11 March) on the Gaia hypothesis came away feeling that their evening of epistemology clarified vague notions more successfully than any of the discussions of plankton, plate tectonics or palaeontology.

The Gaia hypothesis — named after the Greek Earth goddess — is, briefly, that the Earth's climate, geology and life function as a concerted whole. Most participants agreed that what emerged was not a definition of what Gaia amounts to, but the exclusion of several things it does not.

The week began with a rousing defence of Gaia by its originator, James Lovelock, who traced its origins back to the eighteenth century, before science had split irretrievably into the disparate and uncommunicative disciplines of today. He cited the views of James Hutton, one of the founders of modern geology, who declared in 1785 that the proper study of the Earth should be conducted by physiologists, who would view the complete system rather than an assembly of its parts; this followed the discovery of the circulation of the blood, a unifying idea of wide influence.

Since then, Lovelock complains, science has become a set of specialisms, and the Victorian idea of science as the elucidation of particular mechanisms means that those who try to cross boundaries tend to be regarded as dilettantes at best, charlatans and mystics at worst.

Mechanisms

Paradoxically, Gaia has lately gained some attention in respectable circles because specific mechanisms by which its aims might be accomplished have been proposed. Strict devotion to the Gaia hypothesis would, for example, require that if the heat output of the Sun were to fall, organisms on the Earth would respond collectively by producing more carbon dioxide, or other greenhouse gases such as methane, to maintain a habitable temperature.

Although it is indisputable that living

organisms constitute an important element of the chain of geological, chemical and biological processes determining the carbon dioxide content of the atmosphere, the traditional argument would say that trace gases are fixed largely by the geochemistry of the Earth's 4,500-million-year history, and that life has evolved to cope with the conditions thus created. By corollary, as conditions have changed, so life has adapted willily.

What adherents of Gaia have lacked is a demonstration not only that life is capable of influencing its environment in more than minor ways, but also that such influences would have the effect of preserving the *status quo*. A possible mechanism was proposed last year by Charlson, Lovelock, Andreae and Warren (*Nature* 326, 655; 1987).

Many kinds of oceanic phytoplankton produce the gas dimethyl sulphide (DMS) which escapes to the atmosphere and forms a sulphate aerosol. The sulphates are important nuclei for cloud condensation; if cloud cover increases, the Earth's albedo can increase (clouds over oceans contribute significantly to the overall reflectivity of the Earth), and less heat is received at the surface. In principle, if plankton grow faster, the Earth will cool, and the cooling might reduce plankton productivity, thus completing a negative-feedback loop by which the appropriate species of plankton would be the temperature sensor in a global thermostat.

It became abundantly clear at the meeting that nothing as simple as this can be at work. The plankton population does not respond to temperature directly, but depends on the season, on ocean currents, on competition with numerous other species, on the availability of the nutrients for which it is competing and so on. Worse, the emission of DMS into the atmosphere is not proportional to plankton density, but also varies with ocean conditions and the presence of other marine organisms. Half of the Gaian feedback loop is therefore missing, and the other half works erratically.

Philosophical schism hinges on issues such as this. Gaians (to use an abbreviation popular at the meeting) argue that this state of affairs is indeed evidence of the interconnectedness of life on Earth, and that it would be foolish to expect to find a series of isolated and independent mechanisms. Traditionalists, on the other hand, consider this sort of argument an

escape, an indication of the essential untestability (hence scientific poverty) of the Gaia hypothesis, as the response to difficulty is usually a vague appeal to unknown or unknowable complexity.

Much of the conference was taken up with suggestions of different ways in which life and climate might constructively interact, but many of these proposals were merely a sign that Gaians are even more ingenious than Gaia herself.

Take the argument about forest fires. Ecologists accept that fires are, over the long term, essential for vigorous tree growth. Burning clears out undergrowth that might choke saplings, removes old dead wood and recycles nutrients into the soil. Regular fires maintain pine forests in a state of arrested adolescence, warding off an old age in which ground plants take over, trees fail and impoverished soil becomes bog. As an ecological system, a bog is a true dead end, stagnating until some large external force (an ice age, for example) overturns it.

Rejuvenation

May it therefore be the case that forests embody the ingredients of their own rejuvenation by promoting fires at a healthy frequency? Gaians note that many trees give off an inflammable hydrocarbon, isoprene, in sufficient quantities to influence the spread of fire, and that they expend as much as 5 per cent of their energy producing isoprene rather than growing leaves. Why should trees go to such trouble creating a substance for no purpose?

Tree burning inspired remarkably creative thinking. The frequency of forest fires will depend, of course, not only on the properties of trees, but on the oxygen content of the atmosphere. To a Gaian, this dependence is significant. If oxygen were 30 per cent of the atmosphere rather than 21 per cent, burning would be more common, more devastating, and would then be a purely destructive ecological force. But if oxygen were at 12 per cent, fires would be rare and forests would lapse into senescence.

So may not forests be agents of global stability, keeping oxygen at a level that neither suffocates nor consumes life? With more oxygen causing more fires, more ash would be created most of which is eventually washed into the sea, where it increases the phosphorus content. But more phosphorus enhances the rate at

which oxygen dissolved in the ocean is precipitated as mineral sediments on the sea floor. *Ergo*, Gaians argue, the response of forests to more oxygen is ultimately to increase the burial rate of oxygen in the oceans. A stabilizing negative feedback exists.

One can be even more ingenious. If a tree burns completely, all of its carbon ends up in the atmosphere as carbon dioxide. But if another identical tree grows in its place, photosynthesis converts the same quantity of carbon dioxide back into tree material, and releases to the atmosphere precisely the oxygen originally consumed. Over the lifetime of a forest, efficient burning therefore has no overall effect on the oxygen and carbon dioxide levels in the atmosphere.

Charcoal

But trees do not burn completely. Suppose 10 per cent of a burned tree ends up as charcoal lying biologically inert on the ground for longer than it takes a new identical tree to grow. The new tree must get 10 per cent of its carbon from a new source, and in the process releases by photosynthesis 10 per cent more oxygen than the inefficient burning of the old tree consumed. On a medium-term view, long enough for trees to grow back but short enough for charcoal to remain inactive, forest fires paradoxically put oxygen into the atmosphere.

To complete the circle, Gaians observe that, at lower oxygen concentrations, fires burn less efficiently and thus put more oxygen into the atmosphere. This feedback is in fact positive, because the net effect is always to increase oxygen concentrations. But Lovelock emphasizes that a network of positive and negative feedbacks is the essence of a generally stable system. Perhaps the two effects of tree burning, oxygen restoration as the trees grow back and oxygen removal by oceanic phosphorus, combine to stabilize oxygen at 21 per cent of the atmosphere.

Different people become sceptical at different points in this argument. The elementary budgeting of carbon, oxygen and phosphorus clearly omits many sinks and sources, and dynamical considerations (whether climatic or biological) are missing altogether.

But the fundamental debate was not over the details of the reality or efficacy of various feedback mechanisms, but whether even in principle such effects can be construed as evidence for the Gaia hypothesis. Early in the meeting, a distinction between "weak" and "strong" versions of the hypothesis emerged. Where the strong hypothesis maintains, in Lovelock's words, that "the Earth is alive" or "the Earth is a superorganism", the weak version makes the milder assertion simply that mechanisms exist by which a variety of proposed weak Gaia biology

can influence climate, and vice versa, in stabilizing ways.

The business with forest fires is a variety of weak Gaia; the true weakness of weak Gaia is first, that the existence of such mechanisms must be distinguished from more pregnant questions of their origin or purpose, and second, that no number of weak hypotheses demonstrate a strong one. Any weak Gaia hypothesis can be construed by the traditionalists as equivalent to a statement that life is adapted to its environment. A tautology can always be found: trees prosper in 21 per cent oxygen because, if they did not, they would not (a kind of anthropic principle for trees).

But the Earth's oxygen was generated by plants over a long period and, in the interim, life presumably functioned successfully at other oxygen concentrations, the test of success being simply survival. In one view, Gaia, life and climate moved steadily towards the present stable state. In the other, life changed as oxygen was generated, and present conditions are stable perhaps because every available niche is filled, plant life on Earth is producing oxygen at maximum capacity and equilibrium is fixed only as a simple balance of sources and sinks.

Long-term changes in the conditions of life on Earth seem to cause the greatest difficulty, practical and philosophical, for the Gaia hypothesis. If Gaia is supposed to maintain a general equilibrium, do ice ages fall within it, performing a global rejuvenation as forest fires do, or must they be regarded as unpredictable changes pushing Gaia, after readjustment, into a new stable state?

Purposeful evolution

The first view, strong Gaia, inevitably brings with it ideas of collective, purposeful evolution which Lovelock himself finds distasteful; the second, weak Gaia, seems to be little more than a slightly supernatural restatement of the fact that life on Earth has survived, at least so far.

Despite strenuous efforts, it became clear that no scientific question arising out of any study of the history or present state of life on Earth could be made to yield an answer that would distinguish the Gaia hypothesis from more conventional interpretations. This led to a discussion of life on other planets, particularly Mars, which at one time seems to have had flowing water on its surface, and perhaps life; conditions would apparently not have been much more severe than in the dry interior valleys of the Antarctic, where a few hardy organisms survive.

Andrew Watson, long a collaborator of Lovelock's, declared that if life had once flourished on Mars and then died out, the Gaia hypothesis is disproved. But this firm assertion, the nearest at the conference to a testable statement of the strong hypothesis, found little support. And Watson

himself admitted that some 'critical mass' would have to be reached before life could influence climate; a few tiny oases of martian biology would clearly be insufficient. Life could also be extinguished by some catastrophic external agent such as a cometary impact a few orders of magnitude larger than that supposed to have killed the dinosaurs on Earth.

It is even possible to imagine life on Mars being slowly extinguished without upsetting the Gaia hypothesis. Gaia is assumed to be capable of preserving conditions on a planet as long as external factors remain within reasonable bounds. But according to standard theories of stellar evolution, the Sun's brightness has increased by about 30 per cent over the lifetime of the Solar System. On Mars, it may have been that the regulatory power of the martian Gaia was overwhelmed when solar output increased beyond some point, and the system would have collapsed. Not even the strongest of strong Gaia adherents claims that the hypothesis renders life immune from all change.

But there are only three possibilities for life on other planets. It may never develop (almost certainly true of Venus); it could develop and then fade (a remote possibility on Mars); or it could develop and flourish (as on Earth, for the time being). If all outcomes can be accommodated within the Gaia hypothesis, it is difficult to see how the idea can ever be tested.

This philosophical disquiet became the theme of the meeting. The advertised aim, to formulate practical and unequivocal tests of Gaia, was certainly not met. More worryingly, it was impossible to construct even the most far-fetched thought-experiments to do the job instead: we could search for life on every planet in the Milky Way and still not know whether Gaia is correct.

Hard-headed sceptics thus had good reason to leave San Diego feeling that the scientific pretensions of the Gaia hypothesis had been thoroughly dispelled, and that only romantics could persist with it. But (with a few exceptions) they did not. The attack by pure rationalism left Gaians largely unscathed: they unashamedly admit to being romantics and are happy to see the Gaia hypothesis described as metaphor, not theory, because that is what they take it to be. The traditionalists may have achieved their overt aim, of proving Gaia untestable, but the Gaians won a more subversive victory: they provoked a meeting at which people in one discipline talked to people in others, and discovered they had mutual interests.

Lovelock himself was evidently pleased that his efforts led to a small reverse of the tide of fragmentation in science. But now that the Gaia hypothesis has been discussed at a formal scientific meeting, he is looking for more controversial ideas to worry about.

David Lindley

The *ras* oncogene

A structure and some function

Irving S. Sigal

BECAUSE of the prevalence of mutated forms of the *ras* proteins in human tumours, there is intense interest in elucidating the biological function of the three *ras* oncogene proteins — Harvey, Kirsten and N-*ras* — and their normal cellular homologues (see ref. 1 for a review). The membrane-associated *ras* proteins of relative molecular mass 21,000 specifically bind GDP and GTP². Hence, *ras* proteins have been considered as novel G proteins — guanine nucleotide-binding regulatory proteins^{3,4}. The target for the action of *ras* in mammalian cells has been elusive, but there have been two recent and significant advances in elucidating *ras* function and structure. Frank McCormick and colleagues have discovered GAP (GTPase activating protein) in mammalian cells that interacts with the *ras* protein⁵, and Sung-Hou Kim and colleagues have obtained the crystal structure of the normal Harvey *ras* protein complexed to GDP⁶. On page 548 of this issue⁷, Cales *et al.* attempt to relate *ras* structure and function by identifying a region of the protein that might interact with GAP. The pressing issue remains as to whether GAP regulates the *ras* protein, or the *ras* protein regulates GAP.

The trail that led to GAP began with the discovery of oncogenic forms of *ras*, with single amino-acid substitutions (particularly at positions 12 and 61) that transform mammalian cells even at low levels of expression compared with the high levels required by the normal *ras* proteins. There was speculation that the oncogenic substitutions either released *ras* proteins from complexes with an inhibitory protein or altered some key biochemical property of the *ras* protein. Purified *ras* proteins made in *Escherichia coli* possess intrinsic, but slow, GTPase activities that are reduced 10–100-fold by various oncogenic mutations. Oncogenic *ras* was thought to be a G protein gone awry, its reduced GTPase activity permitting the biologically active GTP complex to persist. Some quantitative discrepancies between *in vitro* GTPase activities and biological potencies meant that this model was tested by determining whether more GTP was bound *in vivo* to oncogenic *ras*. Studies with *ras* proteins expressed in yeast cells revealed that, unexpectedly, normal mammalian *ras* exists mostly complexed to GTP rather than GDP⁸.

McCormick and colleagues provided a solution to this paradox. By microinjecting mammalian *ras* proteins complexed with GTP into frog oocytes, they discovered⁵ the cytosolic factor GAP, which

stimulates the GTPase of normal *ras* by 100-fold with no apparent effect on the GTPase of oncogenic *ras*, and appears to be present in mammalian cells but not in yeast cells. The specificity of GAP for normal *ras* can explain the occurrence in mammalian cells of the biologically inactive GDP complex of normal *ras* and the biologically active GTP complex of oncogenic *ras*.

Does GAP function upstream of *ras* as a

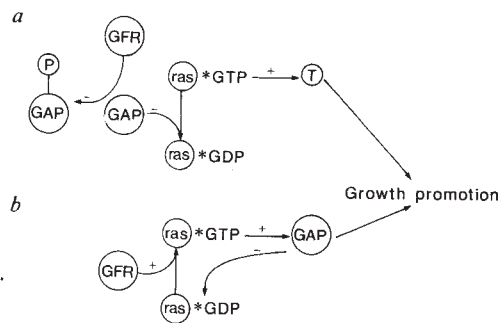
measure the GTPase activities using impure proteins.)

Does this mean that GAP is the target of *ras* and 32–40 is the site for interaction of *ras* with GAP? Not necessarily. The 32–40 region could be the binding site for both GAP and a distinct target protein. It is also possible that the 32–40 region is the site for GAP binding but that substitutions in this region exert effects elsewhere, reducing the affinity of *ras* for its target. If the conformations of the mutants are altered, the 32–40 region might not be a binding site at all.

Fortunately, the spatial locations of critical regions of *ras* are now known from its crystal structure (Fig. 2). The positions for oncogenic substitutions are near the

Fig. 1 Alternative schemes for the actions of *ras* and GAP.

a, In the 'upstream' scheme, GAP is a negative regulator of *ras*. The GTP complex of *ras* interacts with an unknown target T to promote cellular growth. GAP would act as a negative regulator of *ras* by catalysing the conversion of the active GTP complex of normal *ras* to the inactive GDP complex. Growth-factor occupied receptors (GFR), through their phosphorylation of GAP, would promote cellular growth by preventing GAP/*ras* interaction. Oncogenic mutations, by eliminating both the intrinsic and the GAP induced GTPases, would stabilize the active GTP complex of *ras*, circumventing the need for growth factors for cellular growth. **b**, In the 'downstream' scheme, GAP is the target of the active GTP complex of *ras*. According to this scheme, activation of GAP through its interaction with *ras**GTP would lead to cellular proliferation. The conversion of the active GTP complex of *ras* to the inactive GDP complex is concurrent with the activation of GAP. Oncogenic mutations of *ras* would stabilize the *ras**GTP complex and prolong the activation of GAP. GFRs are predicted to facilitate the formation of the active GTP complex of *ras*. The activity of either the target T (upstream scheme) or GAP (downstream scheme) that results in cellular proliferation is unknown.



regulator or downstream of *ras* as its target? If the former is the case, when GAP is brought near *ras* in the cell membrane, it could turn off normal *ras* (Fig. 1a). The *ras* product has been implicated in the signal transduction pathway of oncogene products that are tyrosine kinases and growth-factor receptors, including the insulin receptor (see ref. 9 for a review). If phosphorylation of GAP prevented its interaction with *ras*, then *ras* would persist as the active GTP complex. Alternatively, GAP could be considered as the elusive mammalian downstream target of *ras* (Fig. 1b) based on an analogy with EF-Tu, where the target of EF-Tu, the ribosome, stimulates its GTPase activity¹⁰.

The question of GAP as a target is addressed in this issue by Cales *et al.*, who make use of the amino-acid substitutions near residues 32–40 of oncogenic *ras* that eliminate the transforming activity but do not affect *ras* binding to GTP or its membrane localization¹¹. Cales *et al.* show⁷ that the same substitutions in the normal *ras* protein do not affect the intrinsic GTPase activity but prevent interaction with GAP. (Some caution is warranted in that they

bound GDP; glycine at position 12 contacts the β -phosphate group of GDP, and glutamine at position 61 contacts the glycine at position 12. The structure suggests that the loop comprising amino acids 10–14 is part of the catalytic active site of the GTPase. GAP could function by helping to orient the loop into a more catalytic conformation.

The 32–40 region appears as a coil near the nucleotide-binding site. One might expect that the conformation of this region could change with the binding of GTP. Although it is easy to imagine this exposed region as a site for *ras*–target interaction, some recent experiments raise other possibilities. The *ras* 32–40 region is conserved in R-*ras* which does not transform cells¹². Studies of chimaeric molecules by Lowe *et al.* at Genentech have demonstrated that the carboxy-terminal 80 residues of R-*ras* prevent R-*ras* from having *ras* activity.

Furthermore, Henry Bourne's group at the University of California, San Francisco (personal communication) has discovered, using the chimaeric protein G_i/G_o, consisting of portions of the inhibitory and stimulatory α -subunits of

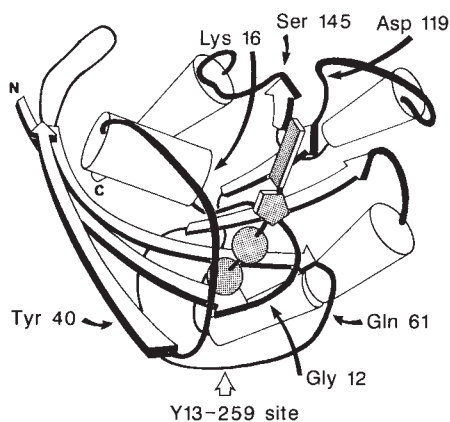


Fig. 2 The backbone structure of normal Harvey *ras* protein complexed to GDP⁶. The structure is that of the amino-terminal 171 residues of 189-residue *ras* products, the protein p21. Residues 10–16 are near the phosphates, residue 30 is near the ribose ring and residues 116–120 and 145–147 form the guanine-base binding site. All the positions where oncogenic substitutions occur lie near the nucleotide-binding site. The epitope for the neutralizing antibody Y13-259, residues 63–73, is shown. The structure is very similar to the previously determined EF-Tu•GDP structure³.

G proteins, that determinants for G_i activation of adenylate cyclase are in the carboxy-terminal portion of the G_i protein. Although analogy with *ras* proteins could be misleading, it seems worthwhile examining regions in the carboxy-terminal domain of *ras* that might be potential sites for protein–protein interaction.

The deletion mutagenesis experiments of Willumsen *et al.*¹³ show that most hydrophilic regions outside the 32–40 region are not involved in *ras* transforming activity. The hydrophobic core itself, and the carboxy-terminal helix that appears to cover it in the GDP complex, remain to be explored. If the 32–40 region is itself not the site for interaction, this region could be involved in transmitting the GTP signal to the carboxy-terminal portion of the protein. When the crystal structure of the GTP complex of the *ras* protein becomes available, comparison of its structure with that of the GDP complex should reveal the exact location of the target binding site, the 'effector region'. □

1. Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1987).
2. Scolnick, E.M., Papageorge, A.G. & Shih, T.Y. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5355–5359 (1979).
3. Bourne, H.R. *Nature* **321**, 814–816 (1986).
4. Casey, P.J. & Gilman, A.G. *J. biol. Chem.* **263**, 2577–2580 (1988).
5. Trahey, M. & McCormick, F. *Science* **238**, 542 (1987).
6. De Vos, A.M. *et al. Science* **239**, 888–893 (1988).
7. Cales, C., Hancock, J.F., Marshall, C.J. & Hall, A. *Nature* **332**, 548–551 (1988).
8. Gibbs, J.B. *et al. J. biol. Chem.* **262**, 10426 (1987).
9. Trahey, M.R. & Jackson, T. *Nature* **328**, 668–669 (1987).
10. Kaziro, Y. *Biochem. biophys. Acta* **505**, 95–127 (1978).
11. Sigal, I.S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4725 (1986).
12. Lowe, D.G. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1015–1019 (1988).
13. Willumsen, B.M. *et al. Molec. cell. Biol.* **6**, 2646 (1986).

Irving S. Sigal is associate director of the Department of Molecular Biology, Merck Sharpe & Dohme Research Laboratories, West Point, Pennsylvania 19486, USA.

X-ray astronomy

Structure of a ring nebula

J.C. Raymond

THEORETICAL models predict that the interaction of a strong stellar wind with the gas surrounding a star produces a thin, dense shell of cool gas filled with very hot, low-density gas of the shock-heated stellar wind^{1,2}. A shell of this sort enclosing supernova 1987A has been inferred from the recent appearance of narrow nitrogen lines in the spectrum of the supernova³. N. Bochkarev reports on page 518 of this issue⁴ the measurement of X-ray emission from the hot interior of the ring nebula NGC 6888 that surrounds a Wolf–Rayet star. He finds that the temperature of the hot gas agrees with model predictions, but that the density does not.

As a massive star uses up its nuclear fuel, it can eject its outer layers to expose a hot core. Fusion has converted the hydrogen in such a star, called a Wolf–Rayet star, to helium. Either nitrogen (in a WN-type star) or carbon (WC-type) is abnormally abundant. The mass ejection which leads to the Wolf–Rayet-type behaviour can take the form of a high-velocity wind, mass transfer onto a binary companion or the ejection of a dense shell at low velocity.

A Wolf–Rayet star emits a very energetic wind, which sweeps the ambient interstellar gas and any stellar gas ejected earlier at low velocity into a thin, dense shell. Photoionization of this shell by the Wolf–Rayet star makes it visible as a ring nebula. NGC 6888 is one of the brightest and best-studied of the ring nebulae. The shell contains about five solar masses of gas expanding at 75 km s⁻¹ away from a WN star⁵. It is enriched in nitrogen and helium, and if one assumes that it is a mixture of normal interstellar gas and gas having the composition of the WN star, then it is composed of 90 per cent interstellar and 10 per cent stellar material⁶.

Theoretical models of such stellar wind bubbles have been widely used to interpret observations of ring nebulae and to model the structure of the interstellar medium. According to the models, the stellar wind expands freely until it encounters a nearly stationary shock front, which heats the gas to a temperature of about 10⁷ K. The pressure of this hot gas drives the dense shell into the surrounding gas.

Current models generally assume that thermal conduction of energy from the hot gas into the dense shell controls the interior temperature and that evaporation of material from the dense shell controls the interior density. Because the temperature and density determine the spectral shape and luminosity of the X-rays

emitted by the hot gas, X-ray observations should be an excellent test of the models⁷. One indication that such a test is needed is that the models predict the kinetic energy of the shell to be 20 per cent of the accumulated kinetic energy of the wind, but the observed kinetic energy of NGC 6888 is only 1 per cent of the estimated kinetic energy deposited by the stellar wind⁸.

Although the faintness of these nebulae makes X-ray observations difficult,

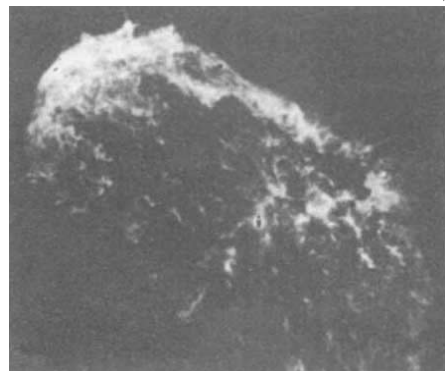


Image of NGC 6888 observed by ionized-nitrogen (N II) emission. The field size is 15 min × 13 min. Logarithmic intensity stretching enhances faint regions. Stars in the field have been suppressed. (Courtesy of J. Hester.)

Bochkarev⁴ uses data from the Einstein satellite to measure the X-ray brightness and spectrum of NGC 6888. The spectral shape is in excellent agreement with model predictions, confirming the role of thermal conduction in determining the interior temperature of the bubble. The X-ray luminosity is an order of magnitude lower than predicted, however, implying that the density of the interior is three or four times lower than expected. Bochkarev suggests that some of the hot gas leaks out of the nebular shell, either because the ambient medium has a porous structure or because of instabilities in the bubble itself^{5,9}. If that is the case, bubble models must be used with great caution. The basic model assumption of pressure balance seems to be violated. □

1. Avedisova, V.S. *Soviet Astr.* **15**, 708–713 (1972).
2. Weaver, R., McCray, R., Castor, J., Shapiro, P. & Moore, R. *Astrophys. J.* **218**, 377–395 (1977).
3. Kirshner, R.P. in *Proc. George Mason Univ. Workshop SN1987A* (ed. Kafatos, N.) (Cambridge University Press, in the press).
4. Bochkarev, N.G. *Nature* **332**, 518–520 (1988).
5. Treffers, R.R. & Chu, Y.H. *Astrophys. J.* **254**, 569–577 (1982).
6. Kwitter, K.B. *Astrophys. J.* **245**, 154–162 (1981).
7. Bochkarev, N.G. *Soviet Astr.* **29**, 509–519 (1985).
8. Lozinskaya, T.A. *Soviet Astr. Lett.* **9**, 247–249 (1981).
9. Hartquist, T.W., Dyson, J.E., Pettini, M. & Smith, L.J. *Mon. Not. R. astr. Soc.* **221**, 715–726 (1986).

J.C. Raymond is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

Palaeontology

Cretaceous unity and diversity

Henry Gee

CRETACEOUS and early Tertiary rocks in the western Sahel region of North Africa display a reassuring continuity from Mali all the way eastwards into Niger. This is one conclusion of a team that has recently returned to Britain from Niger. Such continuity is surprising in view of the complex tectonic and eustatic forces acting on the region throughout the period. The expedition, a joint venture by the British Museum (Natural History) and Kingston Polytechnic, was the culmination of a project begun in 1980 to map the Cretaceous and Palaeogene rocks of Mali.

The sediments in the region reflect three main influences. The main event was the opening of the South Atlantic during the Cretaceous. The team's palaeontological work amplifies the palaeogeographical consequences of continental fragmentation. Lower Cretaceous fossils from the Agadez region of Niger have strong affinities with faunas from other parts of the world, particularly South America, showing that at that time (Aptian-Albian), South America and West Africa were still geographically close. Later African faunas do not show (as far as is known) such strong regional affinities.

At Ingall, in the Agadez region, the team discovered richly fossiliferous floodplain deposits sandwiched between two braided channel deposits. Dinosaur carcasses had been swept into the area by floods and scavenged: the palaeontologists recovered much of the hindquarters of a camarasaurid sauropod (see figure), including both femora, a tibia, most of the pelvis, some vertebrae and a humerus, together with isolated carnosaur teeth. The camarasaurid is believed to be of a type hitherto unknown in Africa, but has been recorded from China and North America.

Other reptiles found at Ingall include fully aquatic crocodiles and a turtle. The fish fauna from Ingall is particularly exciting, as it adds substantially to what is known about African Cretaceous coelacanths and suggests the possibility of correlation with South American faunas. The record until now has been meagre.

Corrigendum

A LITERAL interpretation of the figure and text of N.J. Short's News and Views article "Fraudulent promotion" (*Nature* 331, 393-394; 1988) may have given the impression that the AP-1 transcription factor is activated by the binding of phorbol esters. It should have been made more clear that this is a simplification of a train of events initiated by the binding of phorbol esters to protein kinase C. □

The bone-bed fossils from Ingall can now set into context the bone fragments from Morocco, Algeria, Niger, Egypt and Zaire attributed to the genus *Mawsonia*. The new specimens from Ingall could permit the inter-relationships of the African fossils to be elucidated (a probable consequence being a reduction in species diversity estimation). The species diversity from the Ingall bone-bed has yet to be estimated, but some of the coelacanths are unusually large, possibly up to 1.7 m in length (based on extrapolations from skull

sawfish shark, as well as fin spines of a catfish and remains of the unusual marine teleost *Stratodus*, otherwise known from the slightly older chalk of the Niobrara formation of Kansas (another formation noted for its pterosaurs).

The second event, subsidiary to the development of the Atlantic, is the evolution of a seaway which occasionally stretched across what is now the Sahara from Nigeria to Tunisia. This seaway may have been complete in the Lower Cretaceous, as well as in the Eocene, but the Cretaceous-Tertiary boundary coincides with a major regressive phase. Eocene sediments and foraminiferan faunas in the Gao region of Mali and in Tunisia are remarkably similar, but very little is known about the area in between, so whether the seaway was actually complete

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Uncovering large bones of a sauropod dinosaur during a desert sandstorm. From the fossil expedition to the Sahara, Niger, 1987-88. (Photograph by Phil Crabb, Natural History Museum.)

elements). A coelacanth skull roof is very similar to that of the Aptian-Albian marine coelacanth *Axelrodichthyes*, a recently described genus (Maisey, J.G. *Am. Mus. Novit.* 2866, 1-30; 1986) from the Santana formation of north-east Brazil, well known for its pterosaurs (see last week's News and Views article by David Unwin in *Nature* 332, 398; 1988).

Other fish fossils from Ingall include dentitions and other fragments of the lungfish *Ceratodus*. The toothplates show an unusually large size variation; one is less than 2 mm in diameter. It is unfortunate that lungfish have not yet been found in the Santana formation.

The Upper Cretaceous rocks of the Tahoua area, about 300 km to the southwest of Ingall, are by contrast largely marine, yielding remains of the giant sea lizard *Mosasaurus globudens*, the sea snake *Palaeophis*, a few badly eroded scraps of titanosaurid sauropod and a large variety of fish. The productive sequence at Mount Igdaman produced remains of sharks and rays, including the rostrum of an extinct

in the Eocene remains a matter of speculation. The development of the seaway is linked to the Benue trough, an abortive rift feature in southern Nigeria which is manifested further north as a general sagging.

The third event, superimposing the others, is a confusing eustatic fluctuation. Some transgressive phases are clearly evident, such as the marine Turonian (late Cretaceous) limestones at Abalak, between the Agadez and Tahoua areas of Niger, containing mosasaur fossils.

The entire picture of geological and faunal change is more complex in the mind than on the ground; early Cretaceous transgressions were largely caused by the embryonic Atlantic Ocean, followed by an involved late Cretaceous and Palaeogene choreography of rifting and general sea-level change. But the remarkable sedimentary unity of the area is one which palaeontologists and geologists find particularly satisfying. □

Henry Gee is on the editorial staff of Nature.

Topology

Mysteries of four dimensions

John D.S. Jones

PHYSICS has always provided mathematicians with problems. Physicists are pragmatic people, whose primary interest is in the physical significance of theories rather than in their rigorous justification. This provides mathematicians with one of their traditional sources of good problems. It also creates a little tension: the narrow-minded mathematician dismisses the physicist as sloppy, whereas the narrow-minded physicist complains of mathematicians quibbling over details with no physical significance. Others recognize that the disciplines have different, but compatible, objectives and that each has much to learn from the other. An important example of this flexible attitude is the work of Simon Donaldson, who has used ideas from gauge theory in theoretical physics to prove new results in geometry and topology. His work is published in a series of articles in the *Journal of Differential Geometry*, the most recent of which has just appeared (26, 397–427; 1987).

One of the mathematical advances of the 1980s has been understanding geometry and topology in four dimensions. In mathematics, 'dimension' is used to mean an independent parameter; for example, a space has six dimensions if each of its points is completely determined by six independent numbers. There is no particular reason to restrict the number of dimensions a space should have; spaces with many dimensions occur throughout mathematics. There is, however, something particularly compelling about spaces with three dimensions or even four, where we may regard time as the fourth dimension. A primary aim of topology is to find some respectable classification of these multi-dimensional spaces, rather like the biologist's taxonomic classification; two spaces belong to the same topological class if they have the same basic overall structure, but can differ dramatically in detailed structure.

Classification

Remarkably, the classification is easier in five dimensions and higher: three and four dimensions are by far the most difficult. This does not mean that there is a complete classification in five dimensions or more, but rather that there is a well mapped-out theory that has been adequate to deal with some of the main motivating questions. This is the theory of differential topology that was developed in the 1960s and 1970s. Given the nature of the subject, it is about the best one can hope for. In three dimensions, Bill

Thurston's geometrization programme has generated a great deal of recent activity (see, for example, Ian Stewart's News and Views article in *Nature* 328, 16–17; 1987). What about four dimensions?

Surprisingly theoretical physics provides tools for the classification problem in four dimensions. Gauge theory is considered to be one of the most promising candidates for a theory of elementary particles. The simplest example of a gauge theory is the theory of electromagnetism where the classical (non-quantum) theory leads to Maxwell's equations for the electromagnetic field, and the quantum version is the theory of the photon. In electromagnetic theory there is a phase factor involved, measured by an angle, which corresponds to an internal rotational, or circular, symmetry of the theory.

In general gauge theories, the internal symmetries are more complicated; the classical equations, the analogues of Maxwell's equations, are the Yang–Mills equations and the quantum version is expected to lead to a theory of elementary particles. The problems are to solve the equations and to find the quantum version of the theory. Quite a lot is known about the solutions of the equations, but the quantum version of general gauge theories is still in its infancy. The difficulty is caused by the nonlinearity of the Yang–Mills equations, which is an expression of the complicated internal symmetry of the theory. In electromagnetism, where there is a simple circular symmetry, the equations are linear, solvable and the quantum version is well understood.

Mathematically, the Yang–Mills equations are nonlinear partial differential equations, and traditionally such equations are difficult to handle. There is not much general theory, rather a bag of tricks, and it usually takes great ingenuity to solve individual examples. Much of the initial mathematical work on the Yang–Mills equations was done by Michael Atiyah together with many collaborators (*Fermi lectures: Geometry of Yang–Mills fields*; Scuola Normale Superiore, Pisa, 1979) and there were two things that gave this work an exciting aura. First, there is a direct translation between many of the concepts in gauge theory and concepts in differential geometry; and second, by using a key idea from Roger Penrose's theory of twistors, much of the modern mathematical machinery from differential geometry and algebraic geometry can be used to find solutions.

Donaldson's unique contribution to this circle of ideas was, in some sense, to turn everything on its head. Most of the previous work was along the following lines. Start with a four-dimensional space and then use geometric or topological properties of the space to get information about the solutions of Yang–Mills equations defined over that space. Donaldson argues as follows: if we know something about the solutions of the Yang–Mills equations then we must be able to extract information about the underlying space. This is a remarkable approach, as traditionally the difficult thing is to solve the equations. Donaldson starts by treating the solutions of the equations as, in some sense, the known quantity.

Intersection matrix

What information about the space would we like? There is a basic invariant attached to a four-dimensional space, its intersection matrix. This is a matrix of integers which measures the way two-dimensional spaces intersect. For general reasons this matrix is symmetric — it is unchanged if the entries are reflected in the diagonal. There is a definite notion of algebraic equivalence between matrices such that topologically equivalent spaces have algebraically equivalent intersection matrices. This gives a way of distinguishing between four-dimensional spaces: first calculate their intersection matrices and if these are not algebraically equivalent — something we can decide by a direct manipulation — then the spaces must be topologically distinct. Two fundamental questions arise: does every symmetric matrix of integers arise as the intersection matrix of some four-dimensional space?; and, if two four-dimensional spaces have algebraically equivalent intersection matrices, are they necessarily topologically equivalent?

There are, however, two kinds of space and two notions of equivalence relevant to the theory: topological manifolds and topological equivalence or smooth manifolds and smooth equivalence. In terms of the taxonomic analogy we can regard the purely topological theory as the broad classification and the smooth theory as distinguishing between the finer structure of objects within the same class. If we were discussing five or more dimensions not much harm would be done by this imprecision; there is a definite difference between the purely topological theory and the smooth theory, but it is not too big and it is well understood. In three or fewer dimensions there is no difference whatsoever. In four dimensions however, the difference is large and very poorly understood; indeed it is only because of Donaldson's work that we now know this to be the case.

This point can be illustrated by looking at curves, which are one-dimensional

spaces. A topological curve is just a continuous curve so, for example, it is allowed to have corners; a smooth curve, as the name suggests, is not allowed to have any corners. Given a topological curve, we can 'round off' the corners to produce a smooth curve and up to smooth equivalence the result is independent of the way in which we performed the smoothing. The same is true in two and three dimensions. In five dimensions or more things are different; we can smooth out small enough patches of the space, essentially in only one way, and so the problem becomes one of fitting together these patches to obtain a global smoothing. But this may not be possible and even if it is, there may be different ways of fitting things together. Thus the problem is essentially a combinatorial one. In four dimensions it is still possible to smooth off small enough bits but there is no unique way of doing this: different smoothing processes can lead to vastly different results and this is true no matter how small the piece of the space being treated. This phenomenon is a great mystery. Nothing comparable occurs in any other dimension, indeed precisely the opposite is the case and this is the starting point for the classification theory in dimensions other than four.

Donaldson shows that when working with smooth manifolds and smooth equivalence, the answer to both the fundamental questions above is no and, what is more, no in a very big way. This should be contrasted with earlier results of Michael Freedman (*J. different. Geom.* 17, 357–453; 1982) who showed that when working with topological manifolds and topological equivalence, the answer to the first question is yes, and the answer to the second is that there are at most two spaces with a given intersection matrix. It is the difference between Donaldson's results and Freedman's results which is at the heart of the mysterious features of four dimensions. Donaldson shows how to compute the intersection matrix from the solutions of the Yang–Mills equations. This comes as something of a surprise because, at first sight, there does not seem to be any direct connection. The key ingredient is the existence of instanton solutions to the equations; these are concentrated in a very small ball and they behave like particles placed at the centre of the ball. This gives a way of recovering the points of the space from the solutions of the equations. Donaldson goes on to show how to refine his basic method to describe further, more subtle invariants of smooth manifolds, invariants that can distinguish between smooth manifolds even though they may be topologically equivalent. Again this is a very original line of thought.

What are the implications of Donaldson's work? It shows that the case of

four dimensions is special and that there are many mysteries yet to be unravelled; there are qualitative phenomena which have no counterpart in any other dimension. It also provides tools to measure the complexity of the classification problem in four dimensions. Perhaps the most important next step is to try to see whether, in fact, Donaldson's methods

reveal all of the complexities. To put it another way, are there any more big surprises in store? There is also one important philosophical implication in Donaldson's work which is quite inescapable: that mathematicians need physicists. □

John D.S. Jones is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Atmospheric chemistry

Sources of increased methane

G.I. Pearman and P.J. Fraser

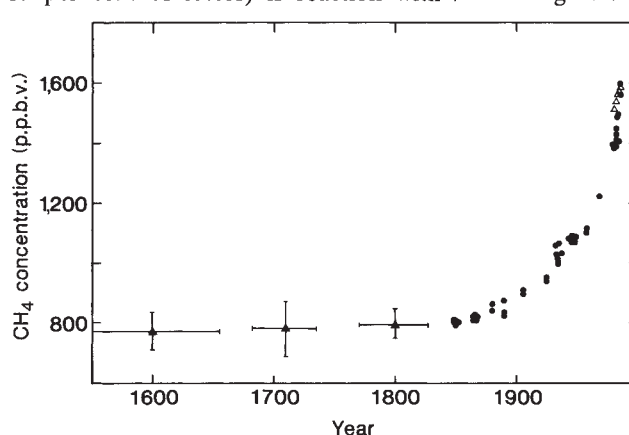
METHANE, like carbon dioxide, is an important atmospheric trace gas that contributes to the greenhouse effect. The increases in the concentration of both since 1800 (150 and 23 per cent, respectively) may have caused significant warming already, if negative feedback mechanisms (such as cloud cover) are ineffective. If the present trends for methane and other trace gases continue, further warming will proceed at almost twice the rate expected to arise from increased carbon dioxide levels alone¹. On page 522 of this issue², Lowe *et al.* present measurements of the radiocarbon (¹⁴C) levels of atmospheric methane, concluding that almost a third comes from fossil, not living, sources, making these more significant than previously thought.

In the past 5 years, high-precision measurements of methane (CH₄) levels have confirmed³ that this gas is increasing in the atmosphere by about 0.8 per cent per year. Studies of air extracted from polar ice cores show that the modern increase is part of a general rise in concentration that parallels human population growth^{4,5}. Concentrations have increased³ from about 650 parts per billion (10⁹) by volume (p.p.b.v.) 200 years ago to the current level of about 1,680 p.p.b.v. (see figure).

The main sink for CH₄ (accounting for 85 per cent of losses) is reaction with

atmospheric hydroxyl (OH) radicals. As the same is true for methylchloroform (CH₃CCl₃), a man-made trace gas whose source strength is known, it is possible to calculate average global OH levels from observations of the distribution and trends of CH₃CCl₃ (refs 6,7). This estimated OH abundance has been used to calculate a sink strength for CH₄ of 400–600 × 10¹² g yr⁻¹ (Tg yr⁻¹). The current global average tropospheric CH₄ concentration³, 1,680 p.p.b.v., is equivalent to a total abundance of 4,300 Tg, and thus the average CH₄ lifetime is about 7–11 years.

The main cause of the increasing concentration is the growth of CH₄ sources⁶, although decreasing levels of OH (caused by increasing concentrations of CH₄ itself and carbon monoxide) could also contribute⁸. The principal sources — enteric fermentation in livestock and insects (80–180 Tg yr⁻¹), rice fields and natural wetlands (40–350 Tg yr⁻¹), biomass burning (20–110 Tg yr⁻¹), land fills (30–70 Tg yr⁻¹) and gas and coal fields (30–90 Tg yr⁻¹) — account for 275–790 Tg yr⁻¹, which encompasses the estimates for the OH–CH₄ sink^{9–11}. Although the estimates are very uncertain, agricultural sources are known to be increasing; both the area under rice cultivation and the cattle population have grown steadily with human population, doubling in the past 40–50 years¹². The



Atmospheric methane variations in the past few centuries. ▲, Summary of previously published data; ●, unpublished data of D. Etheridge (Australian Antarctic Division) and of us; and △, modern observations from the Cape Grim Observatory, Tasmania.

coal-mining source is probably not growing significantly; however the natural-gas source could be, although to date it has not been considered to be a significant cause of increasing atmospheric CH₄.

Fossil CH₄ contains no radiocarbon ¹⁴C, which has a half-life of 5,730 years. The ¹⁴C levels in methane from living organisms, however, are close to those of modern biological material. Lowe *et al.*, in this issue², use this

difference to assess the contribution from these two source groups. Using accelerator mass spectrometry, they measured the ^{14}C content of modern atmospheric CH_4 . After correcting for fractionation effects, and allowing for ^{14}C from nuclear-bomb tests and nuclear-power plants, they suggest that 32 per cent of atmospheric CH_4 comes from fossil sources.

The source strengths listed above give the ratio as 4–33 per cent, depending on whether the upper or lower estimates of fossil and all sources are used. The larger estimates are clearly in better agreement with that of Lowe *et al.*, but would require the sink rate to be much smaller than that deduced from the CH_2Cl_2 -derived OH levels. An alternative solution consistent with both the OH sink and the observed fossil-to-total source ratio is for the fossil sources to total $130\text{--}190\text{ Tg yr}^{-1}$, rather than $30\text{--}90\text{ Tg yr}^{-1}$. This could imply that the losses from coal mining and natural-gas fields are being underestimated or that there are other fossil sources. For example, tundra peat bogs release an estimated $70\text{--}75\text{ Tg yr}^{-1}$ of CH_4 (refs 11, 13). The carbon in tundra bogs could be several thousand years old⁴. If, for example, the radiocarbon age is about 6,000 years, then the fossil source rate would be $70\text{--}130\text{ Tg yr}^{-1}$. Methane hydrates, with a very uncertain release rate of up to 160 Tg yr^{-1} (ref. 11), could contribute to one fossil source.

The work of Lowe *et al.*² improves our understanding of the CH_4 budget and could help explain the increased atmospheric concentration of CH_4 . If it could be measured, the pre-industrial record of $^{14}\text{CH}_4$ trapped in ice cores (about 10 tons of ice need to be sampled) would reveal the contribution of fossil sources to long-term trends. Better estimates of the sources and their isotopic composition will further improve our understanding and lead to more reliable prediction of further greenhouse effects. □

1. Dickinson, R.E. & Cicerone, R.J. *Nature* **319**, 109–115 (1986).
2. Lowe, D.C. *et al.* *Nature* **332**, 522–525 (1988).
3. Steele, L.P. *et al.* *J. Atmos. Chem.* **5**, 125–171 (1987).
4. Rasmussen, R.A. & Khalil, M.A.K. *J. geophys. Res.* **89**, 11599–11605 (1984).
5. Pearman, G.I. *et al.* *Nature* **320**, 248–250 (1986).
6. Fraser, P.J. *et al.* *J. Atmos. Chem.* **4**, 3–42 (1986).
7. Prinn, R. *et al.* *Science* **238**, 945–950 (1987).
8. Thompson, A.M. & Cicerone, R.J. *J. geophys. Res.* **91**, 10853–10864 (1986).
9. Bingemer, H.G., & Crutzen, P.J. *J. geophys. Res.* **92**, 2181–2187 (1987).
10. Seiler, W. & Conrad, R. in *The Geophysiology of Amazonia, Vegetation and Climate Interactions* (ed. Dickerson, R.E.) 133–162 (Wiley, London, 1987).
11. Ehhalt, D.H. in *The Changing Atmosphere*, Dahlem Workshop Report PC7 (eds Rowland, F.S. & Isaksen, I.S.) (Wiley, in the press).
12. WMO/NASA Atmospheric Ozone 1985: World Meteorological Organization Global Ozone Research and Monitoring Project, Report No. 16 (1986).
13. Matthews, E. & Fung, I. *Global biogeochem. Cycles* **1**, 61–68 (1987).
14. Sheppard, J.C. *et al.* *J. geophys. Res.* **87**, 1305–1312 (1982).

G.I. Pearman and P.J. Fraser are in the Division of Atmospheric Research, CSIRO, Station Street, Aspendale, Victoria 3195, Australia.

Developmental neurobiology

The milieu is the message

N. Joan Abbott

As the cells of the embryo divide, grow and migrate, what signals trigger their movements and coordinate their activities? Could the failure of regeneration after damage be caused by lack of correct signals or by a loss of signal receptors? Is signalling disturbed in malignant disease? Are there circumstances in which isolation of a developmental compartment is an advantage? At a recent meeting*, it emerged that patterns detectable in brain development can provide clues to general principles of wider application.

Cellular mechanisms

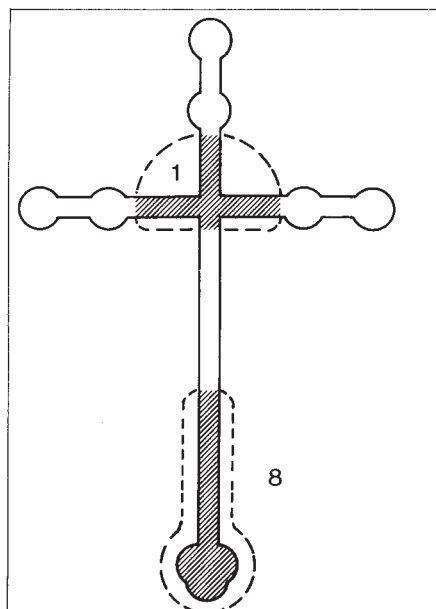
The 'milieu' surrounding a cell is a combination of the fluid medium, glycosylated membrane proteins that project into the medium (the glycocalyx), and polymeric extracellular materials secreted by the cells (extracellular matrix). In tissue culture, many cells start to produce extracellular material as soon as they settle, whereas others will only grow on an appropriate substrate, such as collagen. In most of the adult mammalian central nervous system (CNS), neurons fail to put out new cell processes and axons after damage, partly because of the lack of a suitable substrate¹. The non-neuronal support cells of the nervous system, the glia, are active secretors of extracellular materials. In the peripheral nervous system, the glial (Schwann) cells secrete a tube of basement membrane, condensed extracellular material, which permits axonal outgrowth². And axonal outgrowth in the CNS will occur into a graft of peripheral nerve³, suggesting that peripheral but not central glia can provide the correct substrate.

In the chick embryo, retinal ganglion cells extend axons that grow out into the optic nerve and the tectum in the brain. When these neurons are growing out, their route transiently expresses laminin (J. Cohen, Guy's Hospital, London). Cultured cells from embryonic chick retina can also extend neurites on a laminin substrate, but not onto other basement-membrane components such as collagen or fibronectin.

Laminin, a cruciform glycoprotein of relative molecular mass 900,000, has at least two potential cell-binding sites (see figure). Neurites can extend processes on substrates containing only fragment 8, the globular head and part of the long arm of the molecule, suggesting this is the critical binding site (D. Edgar, Max-Planck Institute, Martinsreid). In the chick, anti-

bodies against the membrane receptor integrin prevent neurite outgrowth on laminin⁴. The drop in responsiveness parallels the decline in neuronal laminin receptors⁵. Here the matrix is the 'message', and neuronal sensitivity depends on the presence of appropriate receptors.

Glial cells also produce diffusible messages that act as mitogenic and growth factors on neurons, on other glial cells and



Localization of cell-binding fragments 1 and 8 in the structure of laminin (from ref. 13).

on vascular endothelial cells (M. Noble, Ludwig Institute, London). Glial tumours show abnormal proliferation: is this caused by increased mitogen production or increase in sensitivity to mitogen⁶? Neuropathological criteria for classification of glial tumours (for example, glioblastoma, astrocytoma and oligodendroglioma) correlate poorly with behaviour *in vitro*, whereas if tumour cells are classified according to whether they express particular antigens, their behaviour is much more predictable⁷. Thus, cells positive for fibronectin secrete an endothelial proliferation factor, whereas cells positive for the neural-tube antigen A4 do not. This may be a key factor in controlling the vascularization of a tumour, itself critical in determining tumour growth rate and metabolism.

The composition of the microenvironment can also control the fate of a cell capable of developing in more than one way. It is known that the 0–2A glial progenitor cell of the mammalian optic nerve produces type 2 astrocytes when cultured in the presence of fetal calf serum,

* The Internal Environment of the Developing Nervous System, Southampton University, 9 January 1988.

and oligodendrocytes in its absence⁸. Sympathetic neurons isolated from the superior cervical ganglion of neonatal rats can be grown in culture, and in the absence of other cells they are predominantly adrenergic (D.D. Potter, Harvard University). When co-cultured with heart muscle cells, they start to produce and release acetylcholine, and this effect can be mimicked by conditioned medium from cultured heart cells. Using antagonists to block both adrenergic and cholinergic receptors, further types of neurotransmitter release can be detected⁹, including serotonin, adenosine and several peptides such as somatostatin and substance P. Of a repertoire of at least fifteen transmitters, individual neurons can release as many as four. The different transmitters require different stimulation frequencies and patterns for their release. The target tissue is inducing expression of particular neurotransmitter combinations, mediated in at least one case by a diffusible molecule. Here the fluid medium is carrying the 'message'. If such plasticity is seen *in vitro*, the picture *in vivo* must be even more complex.

Whole tissue mechanisms

At the single-cell level, the extracellular milieu has important signalling capacity. In tissues where signalling is most complex, and where particular cell-cell interactions are most critical, we might expect to find mechanisms for isolating the tissue milieu from other body compartments to preserve the patterns and gradients of tissue signals. This is exactly what is found in the vertebrate brain, and the compartmentation of the embryonic brain (where connections are developing) is different from that of the adult. The compartmentation gives clues about what features of the milieu need to be controlled during development.

The adult vertebrate brain is separated from the blood compartment by the blood-brain barrier and the blood-cerebrospinal fluid (CSF) barriers, but brain interstitial fluid and CSF are in continuity, and both contain low concentrations of proteins¹⁰. In most groups, the barrier is attributable to tight junctions between vascular endothelial cells in the brain, and epithelial cells in the choroid plexus. In the fetus, the tight junctions are similar in strand number and complexity to those of the adult (K. Møllgård, Panum Institute, Copenhagen), and appear to be as restricting to protein, although alternative routes for protein transfer across the barrier may be present (N.R. Saunders, Southampton University). Fetal CSF is high in protein, and is isolated from the brain interstitial fluid by an unusual single-stranded junction between ependymal cells lining the brain cavities, the ventricles (Møllgård). The source of the CSF protein is partly the plasma, but

Why Lake Michigan is not green

ONE of the abiding questions of ecology is what ultimately determines the biomass of primary producers. Asking "why is the world green?", Slobodkin, Smith and Hairston (*Am. Nat.* 101, 109–124; 1967) outlined some of the possible answers. At one extreme, it could be that predators keep herbivore populations low, so that the abundance of green stuff depends mainly on resource limitations. At the other extreme, plant biomass may be set mainly by the shifting balances in biochemical and other 'arms races' with the creatures that eat plants, with underlying resources being relatively unimportant. Anyone who has compared the teeming abundance of animal life around a coral reef with the relative quiet and vegetative lushness of a tropical rainforest will recognize that there is unlikely to be any single answer.

Many problems of environmental management require an understanding of the factors governing primary production. What, for example, should one do to prevent excessive growth of algae or other photosynthetic organisms in a lake? If the abundance of such organisms depends on nutrient input, efforts should be concentrated on preventing runoff of phosphates and other materials into the lake. But if primary production is regulated by the interactions with herbivores, and ultimately by food-web structure, one must worry about the changes likely to be wrought by the introduction of predatory fish.

As John Lehman observes elsewhere in this issue (*Nature* 332, 537–538; 1988), Lake Michigan is a case in point. Input of phosphates and other nutrients to Lake Michigan has been restricted for some time, with the aim of keeping the lake oligo-

trophic. At the same time, introduced salmon have reduced populations of alewife and other planktivorous fishes. Debate about the long-term effects of this combination of fishery policy and nutrient control is coloured by whether one believes the biomass of primary producers — and thence water quality — is governed mainly by nutrient input or by the structure of the food web.

Lehman provides a careful analysis of a beautiful natural experiment, which seems to answer the question for Lake Michigan. The North American Great Lakes have recently been invaded by a cladoceran predator from the Old World, *Bythotrephes cederstroemii*. Lehman compares the data for the abundance of *Daphnia* species and other zooplankton in 1985 and 1986 (before the advent of the predator) with corresponding data in 1987 (after *Bythotrephes* populations had invaded), and shows that most herbivore populations were significantly depressed. The diminished effects of herbivory, however, appear to have had no effect on the concentration of particulate chlorophyll that Lehman measured in the top 20 metres of the lake at an offshore station; Lehman reasonably assumes that this chlorophyll concentration reflects algal biomass. He thus concludes that "the maximum biomass of algae in Lake Michigan is constrained by forces of other herbivory". Apart from its relevance to a management problem of importance, Lehman's paper is exemplary in its opportunistic use of an unplanned event.

Robert M. May

Robert M. May is Class of 1977 Professor of Zoology at Princeton University, Princeton, New Jersey 08544, USA.

immunocytochemical and molecular-genetic techniques (H. Soreq, Hebrew University) show that brain cells are rich in several of these proteins, such as fetuin, albumin, α -fetoprotein and transferrin, and may be synthesizing them.

One explanation for this compartmentation is that the ependymal zone bordering the CSF (source of dividing and migrating neurons and glia) could need a high-protein medium to stimulate proliferation, whereas the brain interstitium needs a low background protein concentration so that specific proteins produced by neurons (and presumably glia) can act as effective local growth factors, morphogens and signals. The ependymal junctions gradually disappear during development, in parallel with the decline of the CSF protein concentration¹¹. Effective ion regulation of the interstitial fluid and CSF comes in relatively late in development¹², presumably correlating with the maturation of excitability mechanisms and integrated synaptic activity. Thus, cell layers acting as diffusion barriers in the

fetal brain seem to be designed to segregate proteins in different compartments. This suggests that the most important extracellular signals in developing brain are proteins rather than small ions. Although most developing tissues do not show the extreme isolation of compartments seen in brain, they use the same combination of diffusible and matrix-bound components for signalling and controlling differentiation. □

1. Aguayo, A.J. *et al. J. exp. Biol.* **95**, 231–240 (1981).
2. Ide, C. *et al. Brain Res.* **288**, 61–75 (1983).
3. David, S. & Aguayo, A.J. *Science* **214**, 931–933 (1981).
4. Cohen, J. *et al. Devl Biol.* **122**, 407–418 (1987).
5. Edgar, D. & Nurcombe, V.J. *Physiol. Lond.* (in the press).
6. Libermann, T.A. *et al. Nature* **313**, 144–147 (1985).
7. Watkins, B. & Noble, M. in *The Management of Malignant Disease* (ed Thomas, D.G.T.) (Arnold, London, in the press).
8. Raff, M.C. *et al. Nature* **303**, 390–396 (1983).
9. Matsumoto, S.G. *et al. J. Neurosci.* **7**, 380–390 (1987).
10. Bradbury, M.W.B. *The Concept of a Blood-Brain Barrier* (Wiley, Chichester, 1979).
11. Fossan, G. *et al. Devl Brain Res.* **18**, 113–124 (1985).
12. Jones, H.C. & Keep, R.F. *J. Physiol. Lond.* **383**, 441 (1987).
13. Aumailley, M. *et al. J. biol. Chem.* **262**, 11532 (1987).

N. Joan Abbott is in the Department of Physiology, King's College, London WC2R 2LS, UK.

Sewall Wright (1889–1988)

SEWALL Wright, the last of the triumvirate of Fisher, Haldane and Wright who brought about the marriage of darwinism and mendelian genetics, died on 3 March of complications following a fall on the ice when walking near his home in Madison, Wisconsin. He was 98.

Today, when the union of evolution theory and population genetics is a long-established orthodoxy, it is hard to appreciate that the early mendelians engaged in a bitter debate with the darwinians about the mechanism of evolution. The latter held that evolution proceeded by the natural selection of continuously varying traits; the former that large, discrete mutations, obeying Mendel's laws, were the raw material of evolution, and that continuous variation was irrelevant.

Thanks to the work of Fisher, Haldane and Wright, carried out mainly in the period 1920–30, we now see that this argument was based on a misunderstanding. It is true both that mendelizing mutations are the raw material, and that natural selection within populations, often of continuously varying traits, is the main cause of evolution. To show this required that the behaviour of genes in populations be worked out mathematically. Wright's 1931 paper, *Evolution in Mendelian Populations*, was a major contribution to elucidating this problem.

Although it is customary to refer to the three architects of population genetics as if they were engaged in a joint enterprise, in fact they worked independently. This had its drawbacks, as it led to quarrels that may have been enjoyed by the main protagonists, but were confusing for their students. But it had the advantage that three rather different pictures of evolution emerged, each of which may be an aspect of the truth. Wright's particular contribution was to emphasize the importance of population structure. The world does not contain effectively infinite populations, in which any male can mate with any female; instead, populations are divided into many partly isolated groups, or 'demes'. Thus chance events play an important part. In contrast, Fisher held that most populations were so large that chance changes in gene frequency, so-called genetic drift, would be overwhelmed by the effects of selection. Wright displayed remarkable originality — one could almost say idiosyncrasy — in working out the consequences of the demic structure of populations.

His particular view of evolution can best be understood in the light of two features of his early career. First, attempting to apply genetics to animal breeding, he made a detailed study of inheritance in guinea pigs, and second, he studied the breeding of shorthorn cattle. The first of these enterprises persuaded him of the

importance of epistasis — that the effect of genes at one locus depends on what genes are present at other loci. From the shorthorns, he learnt the importance of group structure: in particular, he was struck by the fact that, if a particular herd achieves success, its genes could be spread (mainly by the bulls) through the breed. From these two insights, his 'shifting-balance' theory of evolution emerged.

The theory is as follows. Because of epistasis, evolutionary progress may be impossible in a large random-mating population. Suppose, for example, that the present genotype, at two loci, is *ab*, and that it would be advantageous to evolve to *AB*, but the intermediates *aB* and *Ab* are of lower fitness. A large population cannot make the jump, but the change could occur by chance in a small deme. In a species consisting of many demes, the transition has a good chance of occurring somewhere. If it does, the new genotype can spread through the species, as the

genotype of a successful herd of cattle can spread through the breed.

Wright's shifting-balance theory is still controversial. It is a legacy of the debate between Wright and Fisher that the theory is less popular in Britain than in the United States. But there is no question of the importance of his contributions to population genetics, in particular to the theory of inbreeding, of genetic similarity between relatives, and to the distribution of gene frequencies in populations. Perhaps his most enduring legacy is an image, rather than an algebraic formula, although he left us plenty of the latter. The image is of populations climbing peaks in an adaptive landscape.

Wright continued to work to the end. The last volume of his great textbook, *Evolution and the Genetics of Populations*, was published in 1968, when he was 79 years old. He had a paper in press in the *American Naturalist* when he died.

John Maynard Smith

John Maynard Smith is Professor of Biology University of Sussex, Brighton BN1 9QG, UK.

Particle physics

New phase for an old theory?

R.D. Peccei

COLLISIONS between heavy ions with just sufficient energy to overcome their mutual electrostatic repulsion reveal a series of peculiar phenomena, involving unexplained line structures in the energy spectra of emitted positrons and correlated electron-positron and photon-photon signals. Although many have speculated about the origin of these effects, no satisfactory solution is yet in sight. The most intriguing explanation for the data, suggested recently by three groups^{1–3}, is that the phenomena are the manifestation of a new phase of the theory of electromagnetic interactions, quantum electrodynamics (QED). Because QED is a venerable theory, whose predictions are tested to eight decimal places, this suggestion has been greeted with considerable scepticism by many theorists.

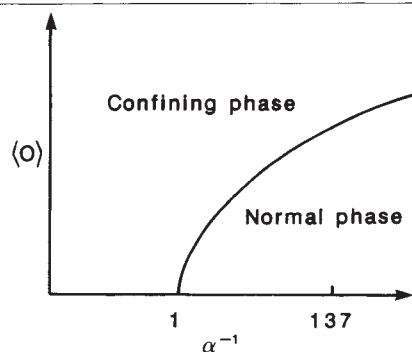
Early experiments at the EPOS^{4,5} and ORANGE^{6,7} collaborations at GSI-Darmstadt revealed a sharp line structure near 350 keV, with a linewidth less than 100 keV, in positron-emission spectra of heavy-ion collisions. The line position is apparently independent of the total charge $Z = Z_1 + Z_2$ of the colliding ions. This feature has been confirmed in further experiments carried out by both collaborations. Various sharp lines appear in the sum-energy spectra, recorded at EPOS^{8,9}, of positrons and electrons emitted back-to-back with kinetic energies near 350 keV. The simplest kinematical explanation is that the correlated signals corres-

pond to the decay of neutral objects, produced at rest in the heavy-ion collision.

With this interpretation, the sum-energy peaks seen in U + Th collisions correspond to states of invariant mass of $1,823 \pm 8$; $1,782 \pm 20$ and $1,630 \pm 8$ keV. A recent experiment at EPOS revealed a line in the sum-energy spectrum of U + Ta collisions corresponding to an invariant mass of $1,770 \pm 8$ keV, and a line seen recently in U + U collisions at ORANGE corresponds to a mass of 1830 keV, agreeing with the EPOS work (Bokemeyer *et al.* and Bederman *et al.* in the *GSI scientific report*, 1987). Also, earlier single-positron spectra produced by Pb + Pb to U + U collisions at ORANGE¹⁰ include lines that seem to be essentially independent of *Z* at about 250, 340 and 410 keV. Finally, correlated pairs of γ -rays emitted back-to-back in U + Th collisions¹¹, correspond to the decay of an object of mass 1,062 keV. It is of particular significance that the correlated peak is extremely sharp, less than 2.5 keV wide, more than a factor of 10 sharper than the correlated electron-positron peaks.

Several aspects of the data fit nicely with the idea that some kind of phase transition has taken place. First, the peaks seem to appear only in the strong, and strongly varying, electromagnetic field of the heavy-ion collisions, as no similar phenomena have been seen in beam-dump searches¹². Second, the narrow width of the signal points to the formation

of soliton-like bubbles of a new phase, whose lifetime is much longer than the characteristic time in which the heavy ions interact strongly (2×10^{-21} s). Indeed, a width of 2 keV corresponds to a period of 2×10^{-19} s, long enough for the ions to separate by 10^3 fm. Third, if the heavy-ion collisions were to trigger the formation of solitonic regions of a new phase of matter, there should be several such excitations, explaining the multiple peaks seen.



Possible phase diagram for QED in the presence of a strong external electromagnetic field. (O), an order parameter for depending on the external field; α , the electromagnetic coupling constant.

The proponents of a new phase of QED use these arguments, and the Z-independence of the observed signals, to justify their hypothesis on phenomenological grounds¹⁻³. They also adduce the mounting theoretical evidence from comprehensive studies of the Schwinger-Dyson equations¹³ and recent lattice calculations¹⁴, that there is a strong-coupling phase of QED. The latter study, in particular, gives clear indications that QED has a strong-coupling phase, in which chiral symmetry is spontaneously broken by the presence of electron-positron condensates. It seems to me, however, that this is something of a red herring, for in the heavy-ion collisions, the coupling constant α (proportional to the strength of interaction) has the usual QED value $1/137$, whereas close coupling occurs when $\alpha \approx 1$. For the new interpretation of the data to be sensible, there must be an order parameter (O) that depends on the external electromagnetic field, such that a phase transition is also possible at weak coupling, as shown in the figure.

Unfortunately, there are some theoretical indications contrary to this picture. Dagotto and Wyld (preprint) recently studied a compact version of QED in a strong Coulomb field on the lattice and found that the phase transition point, rather than moving towards weaker coupling, moves towards stronger coupling. The strong electric field, apparently, tries to dissociate the electron-positron pairs created, impeding the formation of the condensates responsible for the breaking of chiral symmetry, observed at strong coupling. Solà, Wetterich and I reach

similar conclusions (DESY preprint), in the much simpler case of QED in the presence of a space-time constant electromagnetic field. In this case also, even for very strong fields, no breakdown of QED is apparent. The essential point is that nonlinearities, whose growth could drive a transition to a new phase, occur only through the coupling of the external fields to the photons, via an electron loop. Thus, irrespective of the strength of the fields, one must always pay at least a factor of α . These counterexamples notwithstanding, it seems worthwhile to continue exploring whether other external field configurations can drive QED to a new phase. But the phase diagram shown in the figure is clearly an oversimplification.

Theoretical proof that the heavy-ion phenomena are or are not connected with a new phase of QED is unlikely for some time. But phenomenological observations could clarify the matter sooner. The masses and widths of the observed signals should be an intrinsic property of the soliton bubbles of the new phase and therefore should be independent of detailed characteristics of the collision. The rate of production of the bubbles, however, being connected to the strength of an order parameter, should depend on the charge, velocity and scattering angles of the ions. Furthermore, in general, variations in the signal caused by changing some parameter, like Z, should be correlated with those produced by changing some other parameter.

It is important to obtain clear experimental information on these issues. For instance, using the electromagnetic interaction energy as an order parameter gives a production rate for solitons, multiplied by the heavy-ion collision rate, roughly independent of the ions' charge, scattering angle and velocity. Unfortunately, the data from EPOS and ORANGE are contradictory, the EPOS data being particularly sensitive to velocity but not to the charge and the ORANGE data showing just the opposite. Only when these points are settled, and the Z-independence of all peaks is firmly confirmed, are we likely to know more about the possible existence of a new phase of QED. □

1. Celenza, L.S., Mishra, V.K., Shakin, C.M. & Liu, K.F. *Phys. Rev. Lett.* **57**, 55–57 (1986).
2. Caldi, D.G. & Chodos, A. *Phys. Rev. D* **36**, 2876 (1987).
3. Ng, Y.J. & Kikuchi, Y. *Phys. Rev. D* **36**, 2880 (1987).
4. Schweppe, J. *et al. Phys. Rev. Lett.* **51**, 2261 (1983).
5. Cowan, T. *et al. Phys. Rev. Lett.* **54**, 1761 (1985).
6. Clemente, M. *et al. Phys. Lett.* **B137**, 41 (1984).
7. Tseros, H. *et al. Phys. Lett.* **B162**, 273 (1985).
8. Cowan, T. *et al. Phys. Rev. Lett.* **56**, 444 (1986).
9. Bokemeyer, H., Cowan, T. & Greenberg, J.S. in *Physics of Strong Fields* (ed. Greiner, W.) (Plenum, New York, in the press).
10. Koenig, W. *et al. Z. Phys.* **A328**, 129 (1987).
11. Danzmann, K. *et al. Phys. Rev. Lett.* **59**, 1885 (1987).
12. Roridan, E. *et al. Phys. Rev. Lett.* **59**, 755–758 (1987).
13. Fomin, P., Gusynin, V., Miransky, V. & Sitenko, Yu. *Riv. Nuovo Cim.* **6**, 1 (1983).
14. Kogut, J., Dagotto, E. & Kocic, A. *Phys. Rev. Lett.* **60**, 772 (1988).

R.D. Peccei is at the Deutsches Elektronen-Synchrotron DESY, D-2000 Hamburg, FRG.

Daedalus

Salt of the earth

LAST week Daedalus outlined his scheme to develop timber-trees that tolerated heavy metals like copper and nickel in their soil, and accumulated them in their wood. This toxic metal-bearing timber would resist attack by insects and fungi, making it an ideal and long-lasting structural material. Daedalus is now contemplating the same sort of approach to the most deadly phyto-toxic metal of them all — sodium.

Salty soil will not support normal plants. Accordingly, every artificial irrigation scheme is threatened by the inevitable traces of salt in its water. The salt slowly accumulates in the soil and destroys its fertility. This sad fate overtook the irrigated land of the earliest Western civilization in Mesopotamia, and lies in wait for many modern farming regions today.

However, some specialized plants, halophytes, flourish in salty conditions. They do not actually concentrate sodium, but can tolerate high concentrations of it in their tissues. The other noteworthy biological relation to sodium is shown by the mould *Streptomyces*, many varieties of which elaborate ionophore antibiotics. These natural polyethers, reminiscent of the synthetic crown ethers, bind firmly to alkali-metal ions (they kill bacteria by stealing their potassium). So Daedalus plans to combine these separate achievements of biochemistry. By transferring genes from *Streptomyces* to suitable halophyte plants, he hopes to develop a range of plants that can extract sodium even from low concentrations in the soil, and accumulate it in the plant's tissues.

Thus the creeping salinity which sabotages all irrigated land will be held at bay. By growing Daedalus's salt-extracting plants on his land, either at intervals in a crop-rotation scheme, or as a permanent component of his usual crop, a farmer will be able to remove the sodium from his soil as fast as it accumulates.

The extractor-plants will have to be harvested and removed from the land with their saline burden, so Daedalus must make them useful. His first idea was to develop salty varieties of edible crops; salted potatoes for potato-crisps, for example. But his neatest and most ambitious idea is to develop a sodium-fixing plant of the oilseed family. On grinding up and cooking such a plant, its sodium should go to sodium carbonate, its oil should hydrolyse to free fatty acids, and the two should combine to give soap! A soap-plant would be a splendid cash-crop, especially for peasant farmers. And its sodium would be soundly disposed of. It would ultimately go down the sewage system into the sea where it belongs.

David Jones

Has the north-east Atlantic become rougher?

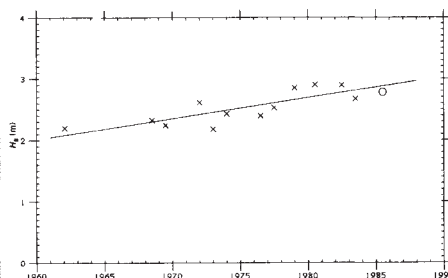
SIR—Measurements of ocean waves made routinely by the Institute of Oceanographic Sciences during the past 25 years suggest that there has been a substantial increase in wave heights in the north-east Atlantic during the period. This trend presents particular problems for those concerned with estimating wave conditions for the design of offshore structures and coastal defences, including estimates of extreme conditions. These estimates are presently obtained by analysing the wave data or values hindcast from wind data assuming that the wave climate is stationary over the years. A better understanding of the climatic variability of ocean waves is needed so that we can assess the likely errors resulting from this assumption of stationarity and eventually take the variability into account in our estimates. Improved knowledge will ultimately be derived from the extensive spatial coverage which will be provided by routine satellite measurements of wave height and of surface wind velocity (together with improved wave hindcast models), but for the immediate future it will only come by continuing wave measurements at specific sites.

Several authors¹⁻³ have concluded that wave conditions over the North Sea and North Atlantic have become more severe in recent years. Their conclusions were based on visual estimates of wave height, on wind and surface pressure measurements and on wave forecast charts.

Instrumental wave records have been obtained by our Institute with few interruptions since January 1962 from a Shipborne Wave Recorder⁴ fitted in the Seven Stones Light Vessel off Land's End. (We believe this to be the longest-running wave recording site anywhere in the world, funded since 1974 by the UK Department of Energy.) Estimates have been obtained from these records of significant wave height (H_s) at three-hourly intervals. H_s is defined as 4σ , where σ^2 is the variance of the sea surface elevation, and is approximately equal to a visual estimate of wave height. Mean values of H_s for each month have been calculated for all months for which at least 90 per cent of the data were successfully recorded, and annual means calculated from these monthly means — not necessarily January to December. The twelve annual mean H_s values obtained to December 1983 are shown plotted against time in the figure (denoted by $\times$). The mean for January–December 1985 is also shown ($\circ$), but 21 per cent of the January data and 16 per cent of the May data for this year are missing, so this value was omitted from the analysis described below.

The twelve annual means were ranked from 1 to 12, from the lowest value of

2.18 m (July 1972–June 1973) to the highest value of 2.90 m (1980). The sum of the ranks for the earlier six years was compared with the distribution of the sum assuming no trend; the Wilcoxon ranking test⁵ rejects the hypothesis of no trend at the 99 per cent level. The straight line shown in the figure was fitted by linear regression. Its slope is 0.034 m yr^{-1} with a standard error of 0.008 — but the regression analysis assumes that annual means are normally distributed, which cannot be strictly true, since $H_s \geq 0$. Applying the Wilcoxon ranking test to monthly



Annual mean values of significant wave height (H_s) at Seven Stones Light Vessel. The value for 1985, denoted by ' $\circ$ ', was calculated from an incomplete data set.

maxima raises problems of the significance of multiple tests, but each of the twelve months (with data to December 1985) has rank sums in the earlier portion of the data less than the expected value assuming no trend.

Modifications have been made over the years to the Shipborne Wave Recorder in the Seven Stones Light Vessel. However, since 1962 the basic measurement technique has not changed and the same method of estimating H_s from the records has been used throughout. We are confident that no instrumental factors can account for the trend in the data.

The 50-year return value of H_s , defined as the value which is exceeded on average once in 50 years, is used for the design of offshore structures in UK waters. At Seven Stones it is about six times the annual mean value of H_s , so a trend of 0.034 m yr^{-1} in the mean indicates an increase in the 50-year return value of about 0.2 m yr^{-1} , from 12 m in 1960 to 18 m in 1990 were the trend to continue.

This Institute has also obtained wave measurements at the north-east Atlantic Ocean Weather Stations (OWS) India ($59^\circ \text{N } 19^\circ \text{W}$) and Juliett ($57.5^\circ \text{N } 20^\circ \text{W}$) since the early 1960s until their withdrawal in 1975. These stations were occupied intermittently by ships of several nations, only a few fitted with a wave recorder, so it took several years to amass what was effectively a 'year's' data of 3-hourly records. However, compound 'years' at both OWS 'I' and 'J' were obtained for the early 1960s and again for the early 1970s. In the 1960s there was no significant difference between the

probability distributions of H_s at the two stations, but by the early 1970s the median H_s at 'J' had increased by 11 per cent and at 'I' by 27 per cent. Wind speeds measured on board the weather ships did not show any such changes, but this is not necessarily inconsistent as waves are an integrated property of the wind over a large area. Meticulous investigations⁶ of all the data revealed no instrumental or analytical reason for these changes; the increased severity of wave height appears to be real.

However, a major problem in determining long-term trends in H_s is the considerable variation from one year to the next. This is seen in the figure, in which the largest change is a drop of 17 per cent from 1971/72 to 1972/73, comparable to the changes at OWS 'I' and 'J'. So it is possible, although unlikely, that the changes at the Ocean Weather Stations could be due to year-to-year variability rather than to a general trend. Clearly, routine measurements are required over many years at a site to determine any long-term trend. However, such measurements can only describe conditions at that site, they cannot differentiate between a widespread increase in severity and, say, a change in storm tracks. These problems will eventually be overcome by the spatial coverage provided by satellite measurements of waves — the radar altimeter gives a good estimate of H_s — and of wind velocities, from scatterometers, which can be fed into improved wave models; but the many years of data required will not be available before the next century. In the meantime we need to continue site measurements.

D.J.T. CARTER
L. DRAPER

*Institute of Oceanographic Sciences,
Deacon Laboratory,
Brook Road, Wormley, Godalming,
Surrey GU8 5UB, UK*

1. Rye, H. *Mar. Sci. Commun.* **2**, 419–448 (1976).
2. Lamb, H.H. & Weiss, I. *Fachl. Mitt. Geophys. BDBw* No. 194 (1979).
3. Neu, H.J.A. *J. geophys. Res.* **89**, 6397–6402 (1984).
4. Tucker, M.J. *Trans. R. Instn. nav. Archit.* **98**, 236–250 (1956).
5. Pearson, E.S. & Hartley, H.O. *Biometrika Tables for Statisticians* Vol. II (Biometrika Trust, London, 1976).
6. Draper, L. *IOS Report* No. 222 (1986).

Assumptions about suicidal behaviour of aphids

SIR—Two recent letters question several assumptions and conclusions from our paper on adaptive suicidal behaviour in aphids¹. These letters reveal several misunderstandings about our study and about evolutionary ecology in general. Thus, we feel a reply is warranted.

First, Tomlinson² argues that our experimental rationale relies on the assumption that our two aphid populations are homogeneous at loci not involved in

anti-predator behaviour. This is incorrect. Our experimental rationale relies instead on the assumption that any direct effects of parasitism of aphid escape behaviour are not significantly different between the two aphid biotypes. This does not require homogeneity at all other loci between the two populations. It is inconceivable, however, that aphids from the dry habitat are simply more susceptible to the direct effects of the parasite. Because of limits on space in our letter we were unable to substantiate the assumption that there were no differences in the direct effects of parasitism. In fact, we have found that aphids from the interior will refrain from suicidal behaviour when their reproductive potential exceeds zero (that is, when parasitized at a later age but assayed at the same age as before)³. Thus, these aphids clearly were not simply more susceptible to direct effects of parasitism than were aphids from the coastal region.

Second, Tomlinson, suggests that aphids from each region should be cross-infected by parasites from the other region to remove possible confounding effects of parasite origin. These experiments have been completed and they provide virtually identical results³. Thus, parasite source cannot explain the different behaviour of coastal and interior aphids.

Third, Tomlinson questions our use of the same on-ground survivorship values for both parasitized and unparasitized aphids. Although we have not measured these values for the former, it is highly probable that they are very similar to the latter, given the extreme temperatures found at ground level in the interior of British Columbia⁴. In fact, one might expect the more stressed parasitized individuals to be even more likely to succumb to environmental stress. Even if, for some remarkable reason, parasitized aphids were able to withstand heat stress to a greater degree than unparasitized individuals, it is unlikely that their survivorship under such circumstances would begin to approach that of aphids remaining on the plant, given the almost twofold difference between these options for unparasitized individuals. Similarly, observations indicate that behaviour following a back up or drop response is not different between parasitized and unparasitized aphids. Thus, we have reason to believe that mortality consequences of the different responses are significantly different between parasitized and unparasitized individuals.

Fourth, we agree that we have not demonstrated that the behaviour is under

genetic control. What we have done, however, is to show that under identical environmental conditions, aphids from two regions respond differently to parasitism. Also, given our lack of knowledge of the genetics underlying these behaviours, it is not clear what the crossing experiments Tomlinson suggested would tell us.

Fifth, Tomlinson makes a series of essentially hand-waving gestures as regards evolutionary stable strategies (ESSs) and the apparent surprising subtlety of the behavioural change. We argue that our original proposition is parsimonious in that it does not require additional unfounded assumptions that the various behaviours have similar payoffs. In fact, the payoffs do differ and each behaviour appears to be used at different frequencies under different conditions, that is, they do not exist in a mixed ESS. With regard to the subtlety of the effect, we remind readers that though such modifications may appear slight in the laboratory, they may be much more pronounced in the field. Also, the results are statistically significant in the predicted direction.

Finally, Tomlinson's argument that asymmetries of selection between hosts and their parasites should favour parasites is unfounded. There are several conditions in which selection on the parasite may not be able to override concurrent selection for suicidal behaviour by hosts: (1) suicidal behaviour is triggered by a complex set of stimuli—this decreases the likelihood that variation in the parasite will be present for traits that could act to subvert this complex behaviour; and (2) the costs of maintaining a host-control mechanism is high relative to the payoff. For example, parasites with exceptionally high fecundity (for instance *Aphidius ervi* ~ 567) may find it exceedingly costly to provide each egg with sufficient neurotoxins to alter the behaviour of every host.

We now turn to Latta's⁵ suggestion that aphids from the coast should drop from plants as frequently as interior aphids because, once off the plant, they are as likely to infect non-kin as kin. If this were the case, we would concur with Latta. In fact, however, it is probably untrue that aphid colonies on 'nearby' plants are more likely to be non-kin, than kin, given the nature of wingless aphid dispersal⁶. Also given the nature of parasite mobility while searching for hosts, it is unlikely that aphid movement to another plant would provide that aphid with sufficient isolation to protect the colony from which it had moved. Isolation is a function of: (1) an aphid's position in space relative to the spatial and genetic structure of aphid populations; and (2) parasite search. We recently developed a mathematical model that explicitly considers the isolation factor and will report on it elsewhere.

Both Tomlinson and Latta question whether the original hypothesis of adap-

tive suicide has been confirmed. We close by reminding scientists that experiments are abstractions that allow us to make inferences about real world events. Nowhere have we argued that we have, in fact, confirmed the host suicide hypothesis but rather, that under a set of well controlled experiments we failed to disprove it.

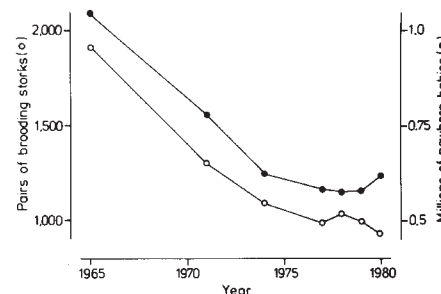
MURDOCH K. McALLISTER

BERNARD D. ROITBERG

*Behavioural Ecology Research Group,
Department of Biosciences,
Simon Fraser University,
Burnaby BC,
V5A 1S6, Canada*

A new parameter for sex education

SIR—There is concern in West Germany over the falling birth rate. The accompanying graph^{1,2} might suggest a solution that every child knows makes sense.



HELMUT SIES

*Institut für Physiologische Chemie I,
Universität Düsseldorf,
Moorenstrasse 5, D-4000 Düsseldorf,
FRG*

1. Fachserie Gebiet und Bevölkerung (Statistisches Bundesamt, Kohlhammer, Stuttgart, 1984).
2. Bauer, S. & Thielcke, G. *Die Vogelwarte* 31, 183–191 (1982).

What are the masses of elementary particles?

SIR—Numerology on the masses of the 'elementary particles' is usually refuted when the experimental results become more accurate, but I now deal with an example that has more than survived the ravages of time and which is to some extent related to the speculations of Eddington¹.

Eddington believed that the fundamental constants could be simply expressed in terms of the integers 4, 6, 10, 16, 120, 136, 256 and 137 which, apart from 137, arise naturally when discussing tensors related to space-time. The prominence of 120 and 136 in Eddington's work was such that he called 136 the "basal multiplicity" and 120 the "number of dormant components in the extended energy tensor". These two numbers can also be described as the numbers of independent real and imaginary elements in a 16 by 16 Hermitian matrix, while 10 and 6 are the

1. McAllister, M. K. & Roitberg, B. D. *Nature* 328, 797–799 (1987).

2. Tomlinson, I. *Nature* 330, 701 (1987).

3. McAllister, M. K. thesis, Simon Fraser Univ. (1987).

4. Roitberg, B. D. & Myers, J. H. *Can. Ent.* 111, 515–519 (1979).

5. Latta, B. *Nature* 330, 701 (1987).

6. Roitberg, B. D., Myers, J. H. & Frazer, B. D. *J. anim. Ecol.* 49, 111–122 (1987).

corresponding numbers for a 4 by 4 matrix. My results depend on the numbers 6, 120 and 136, especially 136.

Eddington introduced the number 137 because he believed, with not much justification, that it was in some sense the true value of α^{-1} , the reciprocal of the fine structure constant. But the present experimental estimate² of α^{-1} is 137.035963 ± 0.000015 , which is not so close to 137 as to suggest that it would be exactly 137 under some appropriate re-interpretation of the experimental results. The closeness of α^{-1} to 136 seems to me to be much more likely to be significant.

For this reason I replace Eddington's $\beta = 137/136$ (Bond's factor) by $\beta' = 1/136\alpha = 1.00761737$ (1 ± 0.000000109). I have also remarked³ that, within experimental error,

$$\frac{m(n) - m(p)}{m(p)} = \frac{1}{720\beta'} = \frac{136\alpha}{720} = \frac{136\alpha}{6 \times 120} \quad (1)$$

where $m(X)$ denotes the rest-mass of a particle X , in this case the neutron (n) and proton (p). The numerator $m(n) - m(p)$ depends only or mainly on electromagnetic energy, so that equation (1), if correct, should eventually shed light on the ratio of the electromagnetic interaction to the strong interaction.

The formula (1) may have been fairly generally overlooked because it was somewhat buried in a discussion in which powers of π were prominent. The number 720 is 'highly composite' in the sense⁴ that it has more factors than any smaller number, so the true reason for its occurrence may well be unrelated to Eddington's speculations. For example, 720 is the order of the symmetric group of order 6, a symmetric group with a property that distinguishes it from all other symmetric groups⁵.

In 1970, equation (1) was correct to one part in at least 11,000, but it is now known, to be correct to one part in at least 51,000 (my work, presented at 1987 meeting of the British Association for the Advancement of Science). The experimental values of the two sides of (1) are now $0.001378398 \pm 0.000000018$ and $0.00137838918 \pm 0.00000000015$, so the discrepancy is only about one half of a standard error.

This degree of accuracy would not be significant if 136 and 720 were arbitrary integers less than 1,000, because there are about a million of such pairs of integers. But since they are by no means arbitrary, I think the result is remarkable. My personal judgment on the basis of the argument up to this point is that the odds are about 'evens' (probability $\frac{1}{2}$) that (1) is 'exact' in the sense of being true to one part in at least a million (not merely one part in at least 51,000). But the odds are

greatly improved by the following observations.

Let X and Y be any pair of elementary particles consisting of quarks and differing in having a u quark in X where there is a d quark in Y or vice versa, and let Y be the heavier. If they are baryons, suppose that they have the same strangeness. Define the positive number

$$R(X, Y) = \frac{m(Y) - m(X)}{m(X)}.$$

It turns out that the current experimental estimates of $R(X, Y)/R(p, n)$ are all close to integers except for the pairs (Ξ^0, Ξ^-) and (B^+, B^0) for which the standard errors are too large to provide appreciable evidence. Moreover all the integers are of the simple form $2^a 3^b$ where a and b are non-negative integers. This is seen from the following table in which all the results are based on ref. 2. (Λ and Σ^0 have the same quark composition but different isospins.) Each integer is the order of a subgroup of the symmetric group of degree 6, and therefore a factor of 720.

Experimental values of $R(X, Y)/R(p, n)$

Quark compositions	X	Y	$R(X, Y)/R(p, n)$	Integer
(uds, uus):	Λ	Σ^+	47.97 ± 0.05	48
(uus, uds):	Σ^+	Σ^0	1.88 ± 0.06	2
(uds, dds):	Σ^0	Σ^-	2.97 ± 0.037	3
(u $\bar{c}$, d $\bar{c}$):	$\bar{D}^0$	D^-	1.83 ± 0.15	2
(u $\bar{s}$, d $\bar{s}$):	K^+	K^0	5.96 ± 0.10	6
(u $\bar{b}$, d $\bar{b}$):	B^+	B^0	0.55 ± 0.47	1
(uss, dss):	Ξ^0	Ξ^-	3.54 ± 0.33	3 or 4

We therefore state as a hypothesis H that $R(X, Y)/R(p, n)$ is an exact integer, with no prime factor exceeding 3, for all pairs of particles as defined previously.

A probabilistic argument based on the table leads to a likelihood ratio or Bayes factor of 100 in favour of the hypothesis H without even allowing for the elegance of the integers. (Details of the calculations can be obtained from the author.) Typically a Bayes factor of as much as 100 corresponds to a tail probability (P -value) of less than $1/1,000$. This large Bayes factor, and the elegant property concerning orders of subgroups, very strongly suggest that the expressions $R(X, Y)$ have deep, though as yet unknown, physical significance. We are therefore justified in believing that (1) expresses a simple relationship between two fundamental constants; that it is correct to one part in at least 51,000, and comfortably consistent with experimental results, is therefore most impressive.

The results can be expressed by saying that, within experimental error, we have

$$\beta' R(p, n) = \frac{1}{720}$$

$$\beta' R(\Lambda, \Sigma^+) = \frac{1}{15}$$

$$\beta' R(\Sigma^+, \Sigma^0) = \frac{1}{360}$$

$$\beta' R(\Sigma^0, \Sigma^-) = \frac{1}{240}$$

$$\beta' R(\bar{D}^0, D^-) = \frac{1}{360}$$

$$\beta' R(K^+, K^0) = \frac{1}{120}$$

$$\beta' R(B^+, B^0) = \frac{1}{720}$$

$$\beta' R(\Xi^0, \Xi^-) = \frac{1}{240} \text{ or } \frac{1}{180}$$

It is natural to conjecture that all of these equations are exact, but even if they are not, it seems to me that a physical explanation is needed for why they are good approximations. It may be noted further that $720\beta' R(\Lambda, \Sigma^+) = 49.98 \pm 0.06$, and might be exactly 50. This is of the form $2^a 3^b 5^c$, and not just $2^a 3^b$, and is not a factor of 720. (Recall that Λ and Σ^0 have the same quark composition so our con-

jecture is not contradicted.) It is also of possible interest that

$$\frac{720\beta'[m(\Sigma^+) - m(\Lambda)]}{m(\Sigma^+)} = 45.00 \pm 0.05$$

and 45 again divides 720. (This time I have divided by the mass of the heavier particle.) It is impossible for this to be exact if $720\beta' R(\Lambda, \Sigma^+)$ is exactly 48 (although the two statements would be mathematically equivalent if β' were replaced by 1), but the occurrence of the number 45 again mildly suggests the possibility that numbers of the form $2^a 3^b 5^c$ might come into a theoretical explanation of our numerical results. The greater simplicity, however, of the numbers $2^a 3^b$ has far more appeal, especially when we think about ratios of the charges of electrons and quarks.

I. J. GOOD

Virginia Polytechnic
Institute and State University,
Blacksburg, Virginia 24061, USA

1. Eddington, A. S. *Fundamental Theory* (Cambridge University Press, 1946).
2. Particle Data Group *Physics Letters* **107B**, 1-350 (1986).
3. Good, I. J. *Physics Letters* **33A**, 383-384 (1970).
4. Ramanujan, S. *Proc. London Math. Soc.* **14**, 347-409 (1915).
5. Burnside, W. *Theory of Groups of Finite Order* 2nd edn (Cambridge University Press, 1911). Reprinted by Dover Publications, New York, 1955.

Heaven in Princeton?

Daniel J. Kevles

Who Got Einstein's Office? Eccentricity and Genius at the Institute for Advanced Study. By Ed Regis. Addison-Wesley, New York: 1987. Pp.316. \$17.95. To be published in Britain by Simon & Schuster on 23 April, £12.95.

THE Institute for Advanced Study, in Princeton, New Jersey, was formally established in 1930 with a gift from Louis Bamberger and his sister, Caroline Bamberger Fuld, and it opened in 1933 with a small staff that included Albert Einstein. The idea for it had originated with its first director, Abraham Flexner, who had recently been studying universities in Britain, Germany and the United States, and who was convinced that scientists and scholars would flourish in an environment in which they were as free as possible from any obligations other than the advancement of knowledge.

Nowadays, the Institute has a permanent faculty of about 25 professors and a corps of visiting researchers, many of them postdoctoral fellows. It is a comfortable place, where the wine cellar holds some 5,000 bottles and gourmet dinners are served twice a week. It includes neither students nor laboratories. "Institute scientists", Ed Regis says, "don't get down in the mud and root around in the dirt of nature".

The Institute, he claims, is a place for theorists — a "Platonic Heaven" — where the main preoccupation of research is something akin to the Platonic Forms, "objects that go profoundly beyond sense experience, and which are apprehensible only by means of thought" (pp. 6–8).

The Platonic Heaven has had its gods — the early staff included not only Einstein but John von Neumann and Kurt Gödel — as well as demigods and mere mortals. But the gods and demigods have all, of course, been human, given to their share of infighting and neuroses (Gödel, in his declining years, came to believe that his food was poisoned and that his doctors wanted to kill him). Reluctance to see these (rather commonplace) features of life in a learned community exposed has, apparently, helped defeat attempts to produce a history of the Institute. An 800-page manuscript was written under the sponsorship of J. Robert Oppenheimer, who was director immediately after the Second World War, but it was suppressed; requests from outsiders to read the manuscript are even now refused (although — so much for secretiveness — a microfilm

copy can be consulted in the Oppenheimer Papers at the Library of Congress). *Tant pis*. We could use a reliable history of the Institute, partly because of the distinction of the people who have worked there, partly because the intellectual vitality of such a place deserves scrutiny and analysis. If you don't produce

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Robert Oppenheimer, director of what he described as an "intellectual hotel". The 'window' (an inset) is the Institute's main building.

at a high level, where there is no obligation except to produce, notes the anthropologist Clifford Geertz, one of the Institute's permanent professors, "it might leave you a little nervous" (p. 281).

Lacking a decent history, we have to make do with Ed Regis's *Who Got Einstein's Office?*. Although both the permanent and temporary faculties include social scientists, historians and philosophers, Regis confines his account to the scientific and mathematical faculty, holding that the Institute exists "to honor the arrogance of mind required of those who are immodest enough to think they can contribute to the task [of comprehending the universe]" (p. xiv). The choice is reasonable, because the Institute is dominated by its mathematicians and scientists, and the security of social science in the Platonic Heaven has been fragile since Flexner's day. (In the mid-1970s, amid much brouhaha, the scientific faculty successfully opposed the appointment of a sociologist from Berkeley who was championed by the then-director, Carl Kaysen, an economist.)

Regis is a member of the philosophy

faculty at Howard University and a contributor to *Omni* magazine, where portions of this book first appeared. The writing is breezy — von Neumann is "Johnny"; Oppenheimer is "Oppie"; astrophysicists are "astros" — and the approach is anecdotal rather than systematic. The book is partly organized around individual scientists, partly around major fields of activity. Chapters are devoted to Einstein, Gödel, von Neumann and Oppenheimer; to astrophysics, cellular automata, fractals and superstrings, among other topics. How representative these treatments are of the scientific and mathematical effort at the Institute is unclear, but they do offer pleasant introductions to their subjects.

Regis provides capable explanations of certain fundamentals of logic and mathematics — for example, Gödel's proof, elementary number theory and how mathematicians go about demonstrating theorems. He is at his best in reporting the observations and explanations of scientists and mathematicians themselves. There is Ron Tubbs on the difference between a merely clever proof and a profound one (the former is just tricky, the latter might open up a whole new branch of mathematics); and Benoit Mandelbrot on why his invention of fractals is important ("many patterns of Nature are so irregular and fragmented, that, compared with *Euclid* —

standard geometry — Nature exhibits not simply a higher degree but an altogether different level of complexity") (pp. 87–90).

There is the astrophysicist Margaret Geller remarking upon data indicating the uneven distribution of galaxies: "When we saw this, it was clear that we were receiving a message from the universe. But the question was, what does it say"; and the mathematical physicist Stephen Wolfram speculating upon cellular automata — the patterns or structures, displayable on a computer screen, that are derived from mathematical rules for self-reproduction from an initial configuration. Wolfram told Regis, apropos a pattern that resembles Pascal's triangle: "... there are examples all over physics and biology of systems that look like that, that grow in exactly that way: crystal growth, for example, cell growth in embryos, the organization of cells in the brain, and so on. The important thing is that the mathematical features of cellular automata are the same mathematical features that are giving rise to complexity in a lot of the world's physical systems" (pp. 179, 244).

However, Regis must be read with considerable caution. He makes small errors (Einstein did not say "God may be subtle, but he is not malicious" in dissent from quantum mechanics but in response to the news, which he heard during his first trip to Princeton, in 1921, that an ether drift had just been detected) (p. 24). And he makes big ones, conflating quantum theory with quantum mechanics and confusing a later interpretation of what the terms in Heisenberg's matrices meant with why Heisenberg resorted to the scheme that became matrix mechanics in the first place (pp. 30–31, 102). Regis permits himself to oversimplify the story of the denial of Oppenheimer's security clearance, declaring that the "conventional explanation" attributes responsibility to a vendetta by Lewis Strauss and that the trustees and the faculty of the Institute joined unanimously to insist that Oppenheimer should remain as director. The truth is that Oppenheimer's fate as a public servant was shaped as much by Edward Teller and big-bomber advocates as it was by Strauss. Moreover, several contemporary letters in the Oswald Veblen Papers from Erwin Panofsky, written at the Institute in 1954, reveal a different reaction by the faculty and trustees. After the security hearing, von Neumann led the mathematicians against an endorsement of Oppenheimer's directorship and they had to be coerced into signing the statement in his support. And although the decision of the trustees to retain Oppenheimer was announced as unanimous, it actually had not been.

Who Got Einstein's Office? is as casual in its judgements of the contemporary Institute as it is in its treatment of historical facts. From the testimony of several scientists — mainly physicists, it seems — Regis finds it difficult to shake off the impression "that the Institute is a nice place to visit but that you wouldn't want to live there, because . . . not enough really happens", adding: "The people there just don't produce" (pp. 278–279). It might be fairer to say that many of the permanent faculty do produce, but perhaps not to a degree consistent with their opportunities. Sections of the Institute do seem to be suffused with an unwarranted aura of self-satisfaction. Still, if the physicists have not shined recently, the addition last spring of Edward Witten, a leading superstring theorist, to the permanent faculty suggests a revival of lustre in the field. The group in theoretical astrophysics has long been active at a high level, partly because of close interaction with its counterpart at Princeton University. Perhaps others are similarly active, too, though one would not know from Regis.

What accounts for the unevenness of output at the Institute is no doubt to be found in its very structure — in, ironically, the lack of any obligation other than to the

advancement of knowledge. For example, not being in the market for students, the Institute lacks one of the key stimuli to academic self-renewal. Useful assessments of the Institute's intellectual consequence would seem to require, at the least, systematic analysis of the research efforts of the permanent faculty, and of the selection procedures for the temporary members and of what they accomplish during their stay, and elucidation of the relationship of the Institute's research activities to those at the neighbouring university.

Who got Einstein's office? It went, first,

to the astronomer Bengt Strömgren, then to the mathematician Arne Beurling, and then to the mathematician Robert Langlands, its current occupant. Regis's book is an impressionistic, not altogether reliable, sketch of the Institute — one that is diverting yet little more arresting than the answer to the question of its title. □

Daniel J. Kevles is the J.O. and Juliette Koepfli Professor of the Humanities at the California Institute of Technology, Pasadena, California 91125, USA. His book The Physicists has recently been issued in paperback by Harvard University Press.

Community spirit

Kenneth Mellanby

Seventy-five Years in Ecology: The British Ecological Society. By John Sheail. Blackwell Scientific: 1987. Pp.301. Hbk £29.50, \$66; pbk £12.50, \$35.

APRIL 12, 1988, is the 75th anniversary of the establishment of the British Ecological Society (BES). To commemorate the event, the society has persuaded John Sheail to write this book, and Blackwell Scientific to publish it.

The term 'ecology' is thought to have been first used by Ernst Haeckel, in 1873, to describe that branch of biology which deals with the interrelationships between living organisms and their environment. Towards the end of the nineteenth century British botanists became increasingly involved in the subject, and in 1904 established the British Vegetation Committee. The committee did excellent work in its restricted field, mainly mapping plant communities. However, there was soon evidence of the need for some organization with wider interests and a larger membership, so the BES, with A.G. Tansley as its first president, came into being.

At first there was hostility from some professional biologists, who viewed the members of the BES as "mere naturalists". The society was not exclusive or elitist, and admitted to membership those with no academic qualifications in the biological sciences, but this did no harm. The naturalists often had an unsurpassed local knowledge of the flora and fauna, which they passed on to the increasing number of professionals who took the leading part in the society's affairs.

John Sheail gives an excellent account of the growth of the BES, and of the increase in the breadth of its interests. At the outset it had only just over a hundred members, but it immediately started publication of *The Journal of Ecology*, the first scientific publication of its kind in the world. At first botanists tended to monopolize the society's activities, but,

largely due to the influence of Charles Elton, zoologists began to play an increasing part, and *The Journal of Animal Ecology* appeared in 1932.

Today the society has some 4,600 members, and it publishes four major scientific journals, a quarterly bulletin and many symposium volumes. It has eight active specialist groups, which hold their own meetings, and has wide international connections. One of its achievements is not recorded by Sheail — thanks to its successful publication policy and to good house-keeping (and because it has avoided the expenses of a costly headquarters building), it is a rich organization, which in good years has added tens of thousands of pounds to its resources. Some of these funds are used to help research, particularly by younger members. But some people are anxious that the BES should make use of its money to become more active in a wide range of topics of public concern.

The public, and the media, have only become aware of ecology in the past 20 years or so. They have not always been well informed, and have paid undue attention to extremists who have little scientific background but who have been accepted as spokesmen for practitioners of the subject. Members of the BES have been concerned; some have wished the society to become more political, others have wished to retreat into their ivory towers. Sheail tactfully records this conflict, though he perhaps underestimates the depth of the feelings of some individuals. Unfortunately, those who have the strongest opinions about environmental topics, in which the voices of ecologists should be heard, may not give the most balanced contribution. Members of the BES must of course be free to express their opinions as informed citizens, but the BES must be cautious when appearing to represent the views of the society as a whole. It has prospered because it has clearly been a scientific, and not a political, organization. □

Kenneth Mellanby, 38 Warkworth Street, Cambridge CB1 1EG, UK, was Director of the Nature Conservancy's Monks Wood Experimental Station from 1961 to 1974.

Sound of distance drums

Vincent Sarich

Molecules and Morphology in Evolution: Conflict or Compromise? Edited by Colin Patterson. Cambridge University Press: 1987. Pp. 229. Hbk £25, \$49.50; pbk £9.95, \$15.95.

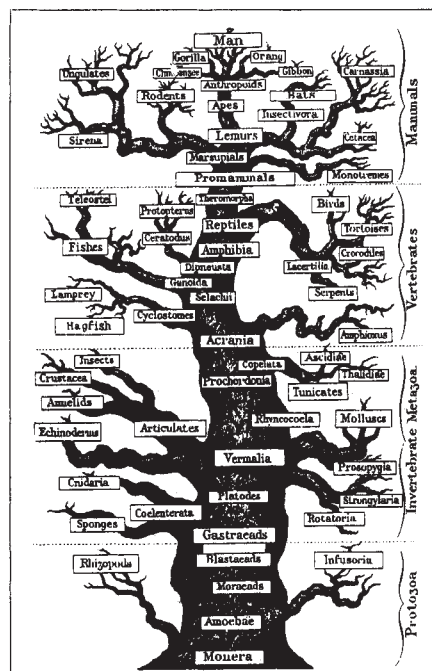
"CONFLICT or compromise?" asks the title of this book. The answer, of course, is "neither" — something only Sibley and Ahlquist, among all the contributors, recognize explicitly. They note, first, that "there is only one true phylogeny" (p. 95), and later that molecules and morphology necessarily complement one another. The molecules are able to "reconstruct the phylogeny with a high degree of accuracy" (p. 118); and, once given the phylogeny, the organismal biologist can use it as a framework within which to interpret structure, physiology and behaviour. We thus want to "cease the debate about 'molecules vs morphology' and begin to co-operate on the basis of molecules and morphology" (pp. 118–119). But that is an ideal which the other authors do not even begin to approach. It is instructive to ask why, some 30 years into the renaissance of the molecular approach to systematics, it is still so often *vs* rather than *and*?

A part of the answer is that non-molecular types can often increase their short-term intellectual fitness by keeping it that way; that is, by not having to accept the constraints that molecular phylogenetic frameworks often impose. The most effective way to do this is to reduce the number and resolving power of such frameworks, and this is readily accomplished by restricting the molecular data worthy of consideration to nucleotide and amino-acid sequences — that is, to 'character-state' data which they consider the only type, morphological or molecular, that can be objectively interpreted within a cladistic framework. This means that distance methods (DNA hybridization, immunological comparisons of proteins) are then rejected on 'theoretical grounds', with the 'theory' being little more than the cladist dogma which states that because distance data are not character-state data, they are not phylogenetically relevant.

It is this explicit rejection of distance methods that is the besetting sin of this book, and it is a sin against science because the rejection is imposed by theory, and not by a demonstration that distance data are indeed phylogenetically irrelevant. If they actually were, their irrelevance could readily be shown by demonstrating the total non-congruence of molecular distance trees with those more or less well-known on the basis of other data. This

would be an impossible task, so it has never been undertaken; and the rejection has simply taken the form of ignoring the distance data because they do not conform to 'theory' — that way the inconvenient point that they work in practice could be ignored.

This is most blatant in McKenna's article on high-level mammalian relationships, but is also, and less excusably, omnipresent in the papers by Andrews, and Goodman and his colleagues. I think that a simple challenge is in order. Is there any relationship among mammals that has been resolved using amino-acid sequences that has not already been resolved by using distance data (generally, immunological



Ernst Haeckel's evolutionary tree of life (1910). The picture is reproduced from Rupert Sheldrake's *The Presence of the Past* (a sequel to *A New Science of Life*, reviewed in *Nature* 294, 32; 1981), which is published today by Collins.

comparisons of various serum proteins)? Indeed, do the available sequence data even give as much resolving power? So why do Goodman and his colleagues persist in ignoring those distance data? Similarly, why does McKenna avoid any mention of the results of my own group on inter-ordinal relationships among the mammals he discusses? Those results, for example, clearly associate sirenians, proboscideans, hyraxes and the aardvark; pangolins and carnivores; sloths, anteaters and armadillos to the exclusion of pangolins and the aardvark; cetaceans and artiodactyls; and indicate very strongly that Glires is not a natural unit. I find it difficult to see all this as anything other than institutionalizing ignorance by concentrating on molecular data that cannot have the requisite resolving power, and ignoring those that have been demonstrated to have it.

Given this institutional bias, it is

incumbent upon those doing good work using distance methods to leave as few targets as possible in their presentations. This Sibley and Ahlquist have failed to do. First, they persist in using a distance statistic, $\Delta T_{50} H$, which is quite arbitrary in that it is not inherent in the melting curves themselves, is overly sensitive to variation in the degrees of hybridization achieved, and allows the ascription of greater precision and range than are actually present in the data. All of this could be avoided by the use of the T_{mode} statistic. The mode of a melting curve, provided there is one, is an integral part of it, and if it is missing that tells us that the relationship involved is beyond the useful range of the approach. Second, they persist in drawing trees which claim too much resolving power. Nodes separated by less than 0.5° or so are simply not credible, and the fact that Sibley and Ahlquist are now claiming rate variation, without allowing for it in the construction of their avian trees, only exacerbates the problem. No method, even in theory, could always give us only dichotomous nodes; thus actual practice, which necessarily involves experimental and interpretative uncertainties, must allow for nodes from which three or more lineages stem. Sibley and Ahlquist, sadly, fail to do so.

Those sins, however, are minor and remediable. Those of a number of other contributors are not. Correction of them is going to require the recognition that the cladistic enterprise of recent years has, in effect, reinvented a flawed wheel, and then found that it rolled with no greater success than it had previously. The reason is that a cladistic analysis tends to be after the fact. That is, one has to know one's organisms very well indeed to know which characters are phylogenetically relevant, and which are not. But how do you judge such utility without having already come to some conclusions as to the phylogeny involved? Simply, you cannot, which makes the delusory possibilities here essentially unbounded. If the answer isn't 'obvious', then by definition there will be characters out there to support each of the various phylogenies. Yet because characters can be neither equal, nor objectively chosen, statistical testing of those phylogenies is precluded.

So where does that leave us? If we follow much of the volume under review, the answer is nowhere. If we do not, and still want to make a practical molecular contribution to systematics, it is to save the sequence approach for situations where distance approaches *do not work*. To do so, we of course need to show that they can work. Given some of the closed minds on view in this volume, that will not be a trivial task. □

Vincent Sarich is a Professor in the Department of Anthropology, University of California, Berkeley, California 94720, USA.

Living around the world

Barry Cox

Biogeography and Plate Tectonics. By J. C. Briggs. *Elsevier*:1987. Pp.204. Dfl. 155, \$75.50.

BRIGGS's book has an exciting title. Study of the interplay between the dispersal of organisms, and the dispersal or collision of continental fragments resulting from plate tectonics, has already solved some long-standing problems and thrown new light on others.

In the past few years, biogeography has also been the forum for an often acrimonious debate between rival groups. First, there are those who believe that dispersal across pre-existing barriers is the important theoretical model for the analysis of historical biogeography. Secondly, there are those who instead support the idea of vicariance biogeography, in which the barrier is believed to have arisen later, dividing a formerly continuous population into two groups with new and independent destinies. Because adherents to the two camps often seem to view their own theory as the only possible one, it is encouraging to read Briggs's statement that his "... book has been written from the point of view that dispersal and vicariance have each played an important role in historical biogeography" — a promise of a balanced approach.

The book is composed of three main parts. The first, "The Northern Continents", is divided into sections that deal with each of the intercontinental connections — the North Atlantic, North Pacific (Beringia), Caribbean and Indo-Australian. In Part 2, on the Southern Continents, Briggs takes a different approach, dealing with each land mass in turn (including New Zealand, Madagascar and India). These differing approaches are sensible, for they reflect the greater complexity of the biogeography of the Southern Hemisphere, resulting from the greater fragmentation of old Gondwanaland.

Within each section, Briggs outlines the apparent historical biogeography of each main group in turn (for example freshwater fish, reptiles, mammals, flowering plants, insects), and discusses or summarizes the plate tectonic data. Although each section does also contain a final summary, it might have been preferable if more integration had been attempted by tracing each land-mass through time and noting the extent to which dispersals demanded by the biotic data are compatible with one another and with the plate tectonic data.

By background Briggs is a specialist on marine zoogeography, and we are reminded of this by the presence of an

unusual third part, on the oceanic plates. Here there are some useful comments on the patterns of speciation and endemism on islands, and on the contrasts between the endemism of terrestrial and marine organisms associated with oceanic islands. The book ends with a bibliography of about 500 entries, and a series of nine world maps from the Triassic to the present day that show the distribution of land and sea. Each map shows only the contemporary coastlines, and is rather roughly and schematically drawn. In general, the book is rather poorly illustrated, for there are only 18 other figures.

At times, I wondered whether Briggs had read widely or deeply enough. Do most palaeobotanists still accept Takhtajan's theory that the flowering plants originated in South-East Asia, rather than the palynological evidence that suggests an origin in the tropics of Africa-South America? Similarly, is the concept of an Arcto-Tertiary Northern Hemisphere

flora still accepted? Is it so certain that New World monkeys and rodents reached South America from the north and not from Africa? Briggs gives us statements, rather than discussions, on such points, so that one can never be sure of the degree of certainty that lies behind each.

Excluding the historical introduction, maps, bibliography and other ancillary material, however, the book is effectively only 135 printed pages long. This is perhaps not enough space to give an authoritative explanation of a subject that requires judgements on the histories and phylogenies of many animal and plant groups. Nevertheless, *Biogeography and Plate Tectonics* is clearly written and well laid out, with an up-to-date and wide bibliography, and will serve as a useful starting point for those who wish to take the subject further. □

Barry Cox is a Professor in the Department of Biology, King's College London (KQC), Campden Hill Road, London W8 7AH, UK.

Progress in gaps

Daniel Goodenough

Cell-to-Cell Communication. Edited by Walmor C. De Mello. *Plenum*:1987. Pp.374. \$59.50 (North America), \$71.40 (elsewhere).

THE 'gap junction' community has evolved to a steady state of having an international meeting once every two years. The availability of a summary of these meetings in book form has been variable, depending on the needs of the meeting sponsors. Nonetheless, we do have one excellent volume containing a review of research on gap-junction-mediated intercellular communication (a 1985 Cold Spring Harbor book, edited by M.V.L. Bennett and D.C. Spray). A forthcoming volume, edited by E.L. Hertzberg and R. Johnson, will include papers presented at a meeting at Asilomar in July 1987, and will be published in July this year by Alan R. Liss.

Cell-to-Cell Communication thus falls temporally between two important books about gap junctions. The title is somewhat pretentious because the volume is almost entirely about gap-junctional intercellular communication. The editor recognizes this, and himself includes a contribution intended to summarize all other forms of intercellular communication — an impossible task for one chapter.

The volume is loosely organized, containing a collection of individual chapters which vary in scope and purpose: they range from those presenting large amounts of the authors' own data (Peracchia, Berdan), to theoretical musings (Sheridan), to concise, critical reviews (Guthrie),

to those that teach the basic concepts (Jaslove and Brink). The book suffers from two unavoidable drawbacks inherent to its species. First, almost all the authors cover only published information, saving their unpublished material for reviewed journals. An exception is the chapter by Berdan, who presents a great deal of eagerly awaited data on progress with the isolation of arthropod gap junctions. The second defect of review volumes — that the contents tend not to be up to date, because of the difficulty of trying to coordinate so many authors — is particularly noticeable here. Research on gap junctions has entered the area of modern molecular biology in the past two years, and this book misses the excitement.

Individually, the chapters present in-depth summaries of published work, and so are useful in that respect. In particular, Guthrie provides a helpful, critical assessment of the literature. But, for example, although Zampighi and Peracchia include all the relevant papers, unlike Guthrie they do not attempt to give any insight as to why different results are found in different laboratories.

Cell-to-Cell Communication thus falls not only between the two meeting reports on gap junctions mentioned at the beginning, but also falls between being an up-to-the-moment summary of data and a tightly edited volume designed for teaching use. All researchers interested in gap-junctional communication will find material of interest here, but most of it has simply been reworked and lacks the excitement of the recent advances in molecular biology. □

Daniel Goodenough is a Professor in the Department of Anatomy and Cellular Biology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.

Response of a general circulation model to a prescribed Antarctic ozone hole

J. T. Kiehl, Byron A. Boville & Bruce P. Briegleb

National Center for Atmospheric Research, PO Box 3,000, Boulder, Colorado 80307, USA

Model simulation of the 'ozone hole' observed in Austral springtime indicates that the ozone depletion leads to a temperature decrease in the lower Antarctic stratosphere of ~ 5 K in mid-October. The temporal evolution of the thermal balance in the control shows that weak upward motion occurs by mid-September and shows the temperature tendency is dominated by the net radiative heating through late September to November. When the ozone hole is imposed, downward motion persists through September to mid-October and the final warming in November is postponed.

PERHAPS the most significant discovery in the atmospheric sciences in the past two years has been the observation of large decreases in Antarctic ozone¹. These depletions in ozone occur in the Austral springtime and since 1979 the severity of the depletion has increased (although not monotonically). In September of 1987 the lowest column ozone amounts ever recorded occurred. Since the discovery of the 'ozone hole', a number of hypotheses have been proposed as an explanation of this phenomenon. The currently proposed hypotheses fall into two categories: those that rely on a chemical mechanism^{2,3}, and those that rely mainly on a dynamical mechanism⁴⁻⁶. There has been much debate as to which of these mechanisms best explains the observations of the ozone hole in the Southern Hemisphere springtime. Our results do not support a major role for the dynamical mechanism.

The present study is concerned with two questions related to these hypotheses. First, does the reduction of ozone observed each spring lead to a temperature change in the Antarctic stratosphere due to the role that ozone plays in the radiative budget? Shine⁷ studied the response of a one-dimensional radiation model to changes in the ozone profile. He found that reductions in ozone similar to those observed did indeed lead to changes in the thermal structure of the lower and middle stratosphere, but with a one-dimensional model he was not able to say whether changes in the dynamics of this region would alter his conclusions. This brings us to the second question to be studied in the present work: what is the relative role of dynamics compared to that of radiative heating in the thermal balance of the middle to lower stratosphere during Austral spring for both the control and ozone hole simulation? The answer to the latter question has great value for determining which mechanism is most important in explaining the occurrence of the ozone hole.

The general circulation model used to consider these two questions is a version of the NCAR Community Climate Model that extends from the surface of the Earth into the upper mesosphere⁸. The model has been run through the Austral springtime using an ozone data set based on observations. Zonal mean ozone is specified as a function of latitude, height and month of year. The control total ozone for the time period under consideration is shown in Fig. 1a.

The model is then run through the same time period, but for an ozone data set that contains the ozone hole. The scenario for the ozone hole is based on the total column ozone results in ref. 9 and the vertical ozone profiles presented in ref. 2. The total column reduction for the experiment is shown in Fig. 1b. The depletion in vertical profile of ozone occurs between 200 and 10 mbar, and decreases linearly away from 90° S to 45° S. The model obtains daily ozone values by linearly interpolating between mid-month tabulations of ozone. The ozone data for the hole experiment was altered for the mid-months of Septem-

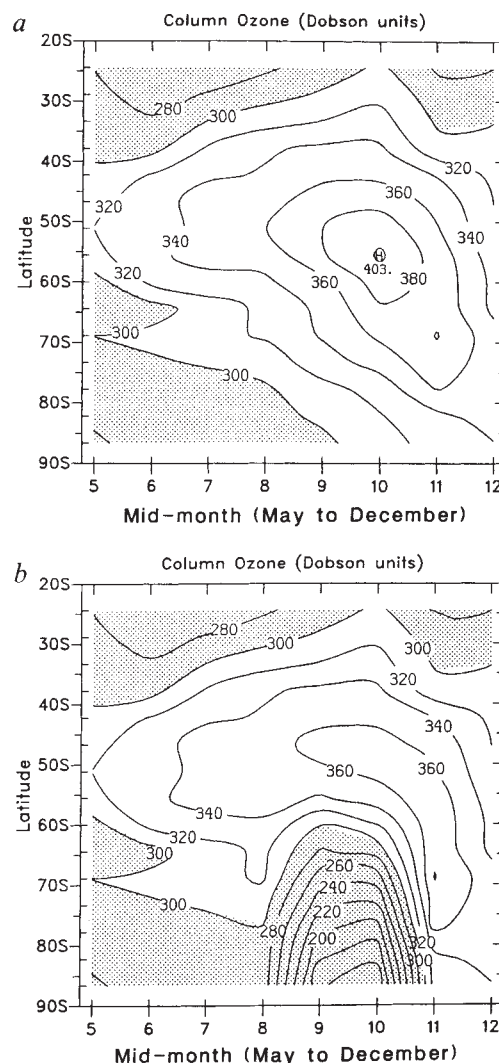


Fig. 1 Total ozone amount for the model control (a) and ozone hole experiment (b).

ber and October. Thus, the hole begins to appear in the model on 16 August and the column ozone amount linearly decreases to a low value in September, followed by a somewhat lower value in mid-October, then increases until 15 November at which time the hole has been removed. Using the control and experiment we address the change in temperature due to an ozone

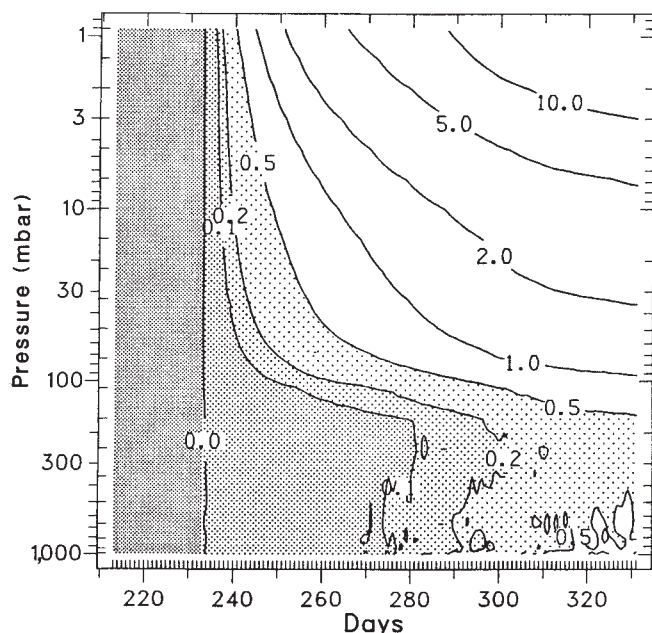


Fig. 2 Zonally averaged solar heating for the control in K day^{-1} at 78°S as a function of pressure and day. Day 213 is 1 August.

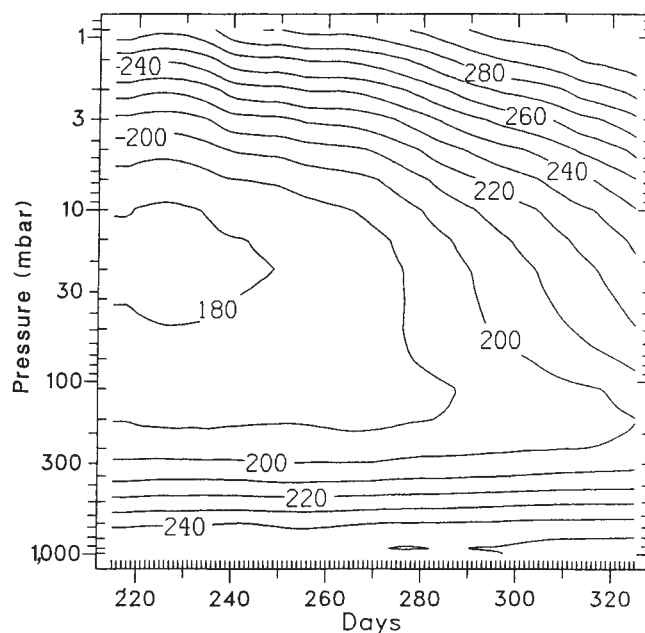


Fig. 3 Temporal evolution of the zonally averaged temperature for the control run at 78°S . Data have been averaged using a 10-day-centred running mean.

hole depletion in Austral spring and consider the temporal evolution of the radiative and dynamical heating terms which contribute to the temperature change. By considering these terms we show the relative role of dynamics versus radiation in determining the evolution of the thermal structure for the control and the experiment.

We do not address the effect of inter-annual variability on our conclusions, but we have performed a second ozone hole experiment using a different initial condition and find no significant changes to our conclusions in the lower stratosphere. Changes in particular dynamical events do result in changes in the temperature evolution in the upper stratosphere (1 to 10 mbar).

Model description and control simulation

The NCAR CCM solves the primitive equations of motion using the spectral transform method. The model truncates all fields at wavenumber 15 which corresponds to a horizontal resolution of $\sim 5^\circ$ of latitude by 8° of longitude. The model contains 26 levels in the vertical (see Fig. 2 for the relative position of the levels). Details of the model can be found in ref. 8. The control simulation was started on 1 May and the hole simulation was started on 1 August. The day numbers correspond to the day of the year (so 1 August is day 213, for example).

An important component of the evolution of thermal structure in Austral spring is the evolution of solar heating at high latitudes⁶. Figure 2 shows the zonally averaged solar heating at 78°S as a function of height and time. The Sun rises on 22 August (day 234) at this latitude. Because of the dependence of absorption pathlengths on zenith angle, the atmosphere above 100 mbar is heated more during the early spring than is the atmosphere below 100 mbar. Even so, solar heating at 70 mbar remains less than 0.5 K day^{-1} through September. The evolution of the net radiative heating (solar plus longwave) is discussed below.

The evolution of the zonally averaged thermal structure at 78°S for the months of August to November for the control simulation is shown in Fig. 3. We show 10-day centred means for all model fields. The temperature evolution of the control is in fair agreement with observations in the lower stratosphere based on a 4-year average of National Meteorological Center

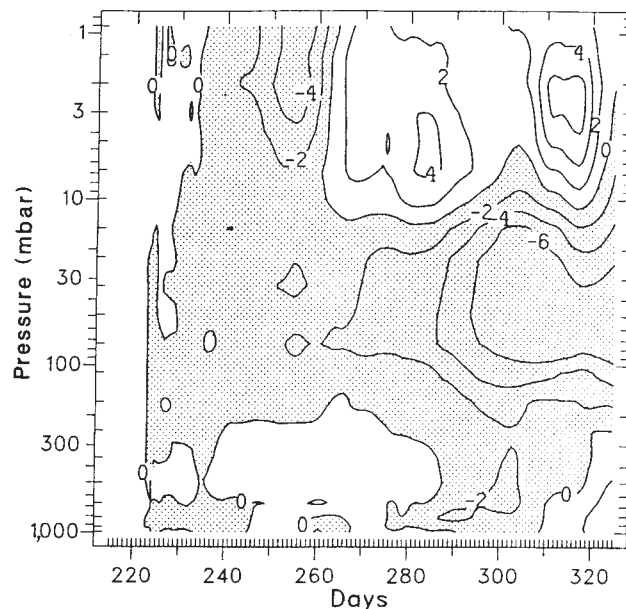


Fig. 4 Temporal evolution of the zonally averaged difference in temperatures between the ozone hole and control simulation at 78°S . Shaded regions signify cooling. Data have been averaged using a 10-day centred running mean.

data¹⁰. For example, the temperature minimum in the model is located at 20 mbar, whereas the observed minimum is near 50 mbar. However, the position of the westerly jet and its eventual reversal in November are reasonably well modelled, although this reversal is delayed by about two weeks.

Temperature response to the ozone hole

The model was then run using the ozone hole scenario described in the previous section. The evolution of the difference in zonal mean temperatures from the control simulation at 78°S is shown

in Fig. 4. Remember that the ozone hole is first felt by the model on 16 August, which is day 228 in the model simulation. The middle to lower stratosphere shows little change at this time, while the upper stratosphere between 1 and 10 mbar begins to experience a cooling. This cannot be ascribed to radiative effects as the hole is small at this time and is located between 10 and 200 mbar. In fact as discussed below, a reduction of ozone in the middle to lower stratosphere should lead to a radiative heating of the upper stratosphere. The cooling of the upper stratosphere is related to a planetary wave event. By the end of September, around day 270, the model is beginning to respond to the decrease in ozone which is largest between 15 September and 15 October. The temperature of the lower stratosphere has decreased by 2 K and continues to decrease through November by ~7 K. The upper stratosphere has now warmed. This warming is mainly due to dynamical heating. This at first seems to be at variance with the finding of Shine⁷, who found a warming just above 30 mbar due to radiative heating, but we have performed radiative calculations with our mid-October ozone hole scenario and find a very small radiative heating (~0.06 K day⁻¹) above 10 mbar. The difference in our result from Shine's arises from the different ozone hole scenarios. Shine's ozone hole profile extends up to just above 30 mbar, whereas our ozone hole extends up to 10 mbar. Thus, very little of the warming in the region between 1 and 10 mbar in our simulation is explained by radiative heating. As pointed out above, the warming in this region is sensitive to the variability of the dynamical heating. The magnitude of the warming, in fact, changed by a factor of two in the companion ozone hole experiment. This large variability in response may be the reason why a warming of the middle to upper stratosphere has not been observed to date. With regards to the cooling of the lower stratosphere, our model's response in mid-October (-5 K) agrees with Shine's radiatively determined response (-6 K) at 20 km (~60 mbar). Temperature decreases of this magnitude should be observable, and indeed there is some indication that a temperature decrease has occurred during the Austral spring over the past seven years¹¹⁻¹³; however, the magnitude of this decrease is uncertain at present.

Roles of dynamical and radiative heating

The zonally averaged thermal balance of the stratosphere is given by

$$\frac{\partial T}{\partial t} = Q_r + Q_d \quad (1)$$

where $\partial T/\partial t$ is the temperature tendency, Q_r is the net radiative heating, and Q_d is the dynamical heating. We now wish to consider the temporal evolution of each of these terms for the control and the ozone hole simulation. The region of interest for the ozone hole problem is near 70 mbar, where in fact the model simulation agrees well with observations. We therefore choose to look at these terms at one level of the model, 74 mbar and again at latitude 78° S. We use a 10-day centred average of all quantities. This averaging period is sufficiently long to eliminate much of the high frequency variability in the temperature tendency term and the dynamical heating, but still short enough to capture the important temporal variations in these quantities. We do not explicitly calculate the dynamical heating term in (1). Instead we take the temperature tendency term and net radiative heating available from the model history tapes and obtain Q_d from (1).

We first compare the net radiative heating for the control and ozone hole simulation in Fig. 5. This figure indicates that ~10 days after the Sun rises the net radiative contribution switches from one of net cooling to net heating. In fact, the pre-sunrise net radiative cooling is 0.15 K day⁻¹, which is not negligible. Thus, the pre-sunrise stratosphere is not near radiative equilibrium, as has been assumed in some dynamical mechanisms arguments^{4,5}. The net radiative heating steadily increases after

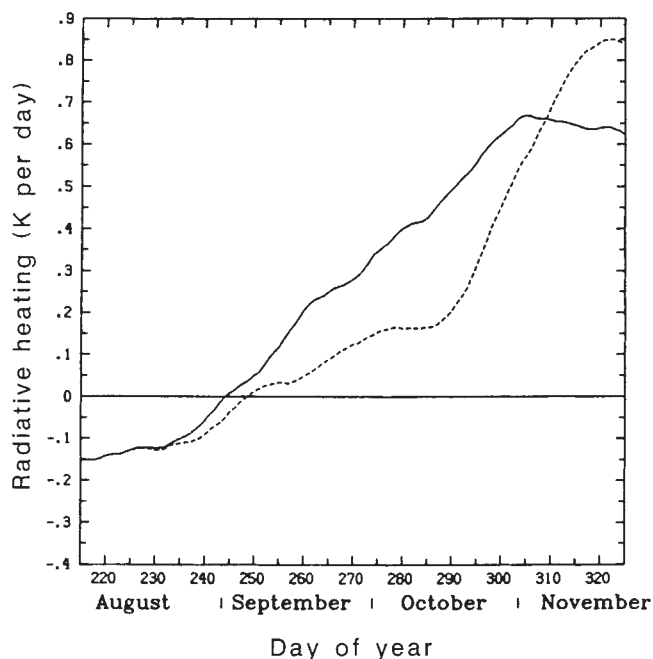


Fig. 5 Temporal evolution of the zonally averaged net radiative heating (K day⁻¹) at 78° S and 74 mbar. Data have been averaged using a 10-day centred running mean. Unbroken line, control; dashed line, hole.

sunrise until the beginning of November, when it begins to decline. This decline is due to the increasing importance of longwave cooling. As the vortex slowly erodes, the increase in stratospheric temperatures leads to increasing longwave cooling. Eventually the rate of increase in cooling begins to overtake the rate of increase in solar heating, and the net radiative heating decreases. The net radiative heating for the ozone hole takes a longer time to switch from cooling to heating due to the reduction in heating caused by the hole. The hole also causes a significant reduction in the rate of increase in net radiative heating through September and half of October. In mid- to late October, the increase in radiative heating is actually larger due to the return of ozone solar heating and smaller longwave cooling resulting from colder temperatures. These colder temperatures prevent the increasing longwave cooling from overtaking the increase in solar heating as early as in the control. It is for this reason that we do not see a decrease in the net radiative heating until the end of November.

All three terms in equation (1) for the control simulation are shown in Fig. 6a. The general behaviour of the dynamical heating, which is dominated by adiabatic compression/expansion due to vertical motions, indicates that downward motion ($Q_d > 0$) predominates before mid-September. After the middle of September weak upward motion prevails ($Q_d < 0$) until near the end of November. Near the middle of November the model experiences a large final warming. The occurrence of such a final warming in November is consistent with observations^{11,14}. The temperature tendency is closely correlated with the dynamical heating term from August to the middle of September. However, after mid-September the temperature tendency follows the net radiative heating until the middle of November, when the final warming leads to a large increase in temperature. We thus find that within a few days after the initiation of net positive radiative heating, the major portion of the radiative heating goes into the temperature tendency and is not balanced by dynamical cooling.

It is important to consider the magnitude of the heating terms in relation to what is required for a dynamical explanation of

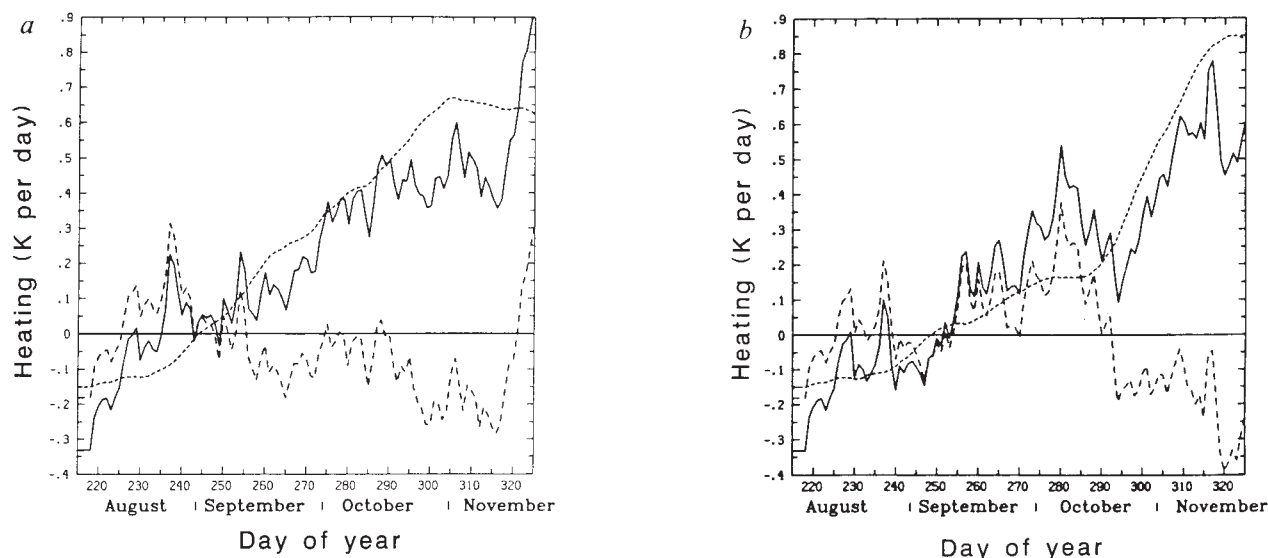


Fig. 6 Temporal evolution of the three terms in the thermal balance equation (1) for 78°S and 74 mbar for *a*, the control and *b*, the ozone hole simulation. Data have been averaged using a 10 day-centred running mean. Unbroken line, temperature tendency; small dashes, net radiative heating Q_r ; long dashes, dynamical heating Q_d .

the ozone hole^{4,6}. The dynamical studies suggest that a heating of at least 0.5 K day⁻¹ is needed for the upward motion to cause a reduction in ozone consistent with observations. This is five times the heating that is present in mid-September in the control simulation. The ozone hole will reduce the net radiative heating of this region, causing an even greater deficit between the actual heating and that which would be required to produce the hole. To consider this question we now turn to the thermal balance of the ozone hole simulation.

Figure 6*b* shows the three terms in equation (1) for the ozone hole simulation. The presence of the hole has significantly altered the dynamical heating term. Downward motion ($Q_d < 0$) now persists through September into the middle of October, suggesting that a negative feedback exists between upward motion and ozone reduction in the lower stratosphere. During this time the dynamical heating is strongly correlated with the temperature tendency. It is only after the middle of October that the temperature tendency is dominated by the net radiative heating term.

Another interesting feature of the hole simulation is the absence of a final warming in November. The colder, longer-lived vortex, due to the imposed hole, has led to a delay in the final warming. This is intriguing because observational data¹⁵ suggest that southern hemisphere final warmings have been occurring later since 1979. The model simulation suggests that it is the reduction in ozone that has caused this to occur, and not a reduction in planetary wave activity that has led to both the existence of the ozone hole and the delay of the final warming. Finally the results of Fig. 6*b* indicate that heating due to Polar Stratospheric Clouds (currently estimated at ~0.1 K day⁻¹; V. Ramaswamy, personal communication) cannot sustain upward motion strong enough to produce an ozone hole in September.

Conclusions

The present study has used a general circulation model to answer two specific questions related to the ozone hole problem. First, how do stratospheric temperatures respond to ozone reductions consistent with the observations of the ozone hole in Austral spring? We find that in the mid to lower stratosphere (10 to 200 mbar) temperatures decrease by as much as 7 to 8 K. The decreases in temperature at 74 mbar extend as far equatorward as 50°S. The largest temperature decreases occur in early to mid-November near the pole. In the upper stratosphere (1 to 10 mbar), a temperature increase occurs with a maximum amplitude of 6 K. This temperature increase is due to dynamical heating, although a smaller warming due to ozone reduction could be expected from changes in the net radiative heating. Second, what are the relative roles of radiative and dynamical heating in determining the thermal balance throughout Austral spring? In the control simulation we find that, when the sun rises, the net radiative tendency takes 10 days to switch over to net heating. During this time much of the local temperature tendency is determined by the dynamical heating, but by mid-September the temperature tendency is dominated by net radiative heating. For the ozone hole simulation downward motion persists through September into mid-October, indicating that the presence of the hole suppresses the evolution to upward motion in mid-September found in the control. Thus a negative feedback between upward motion and ozone reduction exists. These findings indicate that a dynamical explanation of the ozone hole does not seem plausible.

We thank R. R. Garcia for discussion and J. D. Mahlman and K. P. Shine for reviewing the manuscript. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Received 28 December 1987; accepted 26 February 1988.

- Farman, J. C. *et al.* *Nature* **315**, 207–210 (1985).
- Solomon, S. *et al.* *Nature* **321**, 755–758 (1986).
- Crutzen, P. J. & Arnold, F. *Nature* **324**, 651 (1986).
- Tung, K. K. *Geophys. Res. Lett.* **13**, 1308–1311 (1986).
- Tung, K. K. *et al.* *Nature* **333**, 811–814 (1986).
- Mahlman, J. D. & Fels, S. B. *Geophys. Res. Lett.* **13**, 1316–1319 (1986).
- Shine, K. P. *Geophys. Res. Lett.* **13**, 1331–1334 (1986).

- Boville, B. A. & Randel, W. J. *J. atmos. Sci.* **43**, 3015–3034 (1986).
- Stolarski, R. S. & Schoeberl, M. R. *Geophys. Res. Lett.* **13**, 1210–1212 (1986).
- Farrara, J. D. & Mechoso, C. R. *Geophys. Res. Lett.* **13**, 1232–1235 (1986).
- Newman, P. A. & Schoeberl, M. R. *Geophys. Res. Lett.* **13**, 1206–1209 (1986).
- Chubachi, S. *Geophys. Res. Lett.* **13**, 1221–1223 (1986).
- Rosenfield, J. E. & Schoeberl, M. R. *Geophys. Res. Lett.* **13**, 1339–1342 (1986).
- Newman, P. A. *Geophys. Res. Lett.* **13**, 1128–1231 (1986).
- Newman, P. A. & Randel, W. J. *J. geophys. Res.* (submitted 1987).

Gas compression and jet formation in cavities collapsed by a shock wave

J. P. Dear, J. E. Field & A. J. Walton

Physics and Chemistry of Solids, Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

A range of techniques, including high-speed photography and image intensification, are used to study the collapse of cavities by shock waves. Of particular interest are the shape changes of the cavity and the production of luminescence at 'hot spots' in the gas. The relevance of the research to cavitation damage, explosive initiation, and medical uses of acoustic waves is discussed.

THIS article describes an experimental technique for studying the properties and behaviour of gas-filled cavities collapsed by means of a shock wave. The properties studied are the temporal development of the cavity shape and the emission of light from the collapsing cavity. We have used a range of techniques to follow the details of the event, and in particular to record the positions at which 'hot spots' develop. A two-dimensional geometry was used to observe the processes occurring inside the cavity. Single cavities and arrays of cavities were formed in a gelatine layer which was placed between two thick glass blocks, and the cavity was then collapsed by a shock wave produced by an impacting projectile. A major advantage of this technique is that cavity size, shape and number can be controlled, as well as the shape of the shock wave. The interaction of the shock with the cavity and the cavity collapse were recorded by high-speed photography at microsecond framing intervals, with the shocks visualized by Schlieren optics. Finally, an image intensifier was used to record the position of 'hot spots' where the gas luminesces.

The response of a gas space to an increasing pressure field, a pulse of waves or a shock, is important in a range of situations: for example in cavitation damage, in producing 'hot spots' that cause ignition or assist the growth of fast reaction in explosives, in noise generation from flowing liquids, and in medical situations where ultrasonic waves are used for visualization or treatment.

Since the classic work of Rayleigh¹, the collapse of a cavity has been treated with various degrees of sophistication. Rayleigh analysed the collapse of an isolated, empty spherical cavity in a liquid under hydrostatic pressure. Assuming incompressible and inviscid behaviour of the liquid, his analysis gives the time to collapse to reasonable accuracy, but predicts a collapse velocity (and hence liquid pressure) that tends to infinity as the cavity radius approaches zero. In practice, a cavity cannot be considered as empty during the final stages of collapse, and thermal effects (adiabatic heating of the gas, heat of condensation) and liquid compressibility need to be considered. Other workers have included the effects of surface tension and viscosity (for reviews, see refs 2 and 3). The result is that the cavity collapses and then, assisted by any vapour or gas inside, rebounds violently. These processes can produce pressures as high as 1 GPa (10 kbar), which clearly have damage potential, but the pressures fall off very rapidly within a few bubble radii⁴.

Kornfield and Suvorov⁵ were the first to suggest that cavities might collapse asymmetrically and produce a jet. The jet is formed by involution of one side of the cavity and passes across the cavity, striking and penetrating the far surface. If this asymmetric collapse occurs near a solid surface, then there can be a liquid/solid impact with the generation of a 'water hammer' pressure given by

$$P = V\rho_1 C_1 \rho_2 C_2 / (\rho_1 C_1 + \rho_2 C_2) \quad (1)$$

where V is the impact velocity of the jet and ρ_1 , ρ_2 and C_1 , C_2

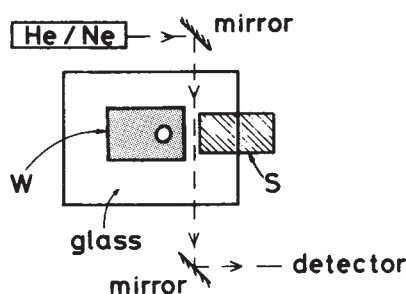


Fig. 1 Impact geometry: W is the liquid/gelatine containing the cavity and S is the impacting striker.

are the densities and shock-wave velocities of the water and solid respectively. The duration of this high pressure is given by

$$\tau = r/C_1 \quad (2)$$

where r is the jet tip radius. For a cavity wall well removed from the solid surface, the impact of the jet on the opposite wall produces a pressure pulse in the liquid of magnitude $\rho CV/2$, which can subsequently interact with the solid.

Two situations arise that can lead to jetting. The first is where jet formation takes place in an asymmetrical pressure field as produced near a solid surface, and has been examined by Plesset and Chapman⁶. The second is when a shock passes over a cavity, causing a jet in the direction of the shock. For the situation analysed in ref. 6, the jet formation is most pronounced for cavities closest to the solid surface. Experiments using laser-induced cavities formed at various distances from the solid surface have confirmed these predictions⁷. Jet velocities, V_j , are typically 130 m s^{-1} for a pressure difference, ΔP , of 1 bar and vary as $\Delta P^{1/2}$. For the case of shock-wave collapse of a cavity, the conditions are somewhat different, in that initially the side of the cavity farthest from the shock is not aware of the collapse process. Consequently, the dynamics of the collapse and the pressure profile around the cavity surface differ from the Plesset and Chapman studies. Numerical codes have been used to model this process^{8,9}, and others (see, for example, ref. 10) have considered the shock-wave collapse analytically. The tip of the jet in the shock collapse model travels at a velocity of

$$V_j \approx 2V \quad (3)$$

where V is the particle velocity of the shock (this is the classic case of a shock reflecting at a free surface and giving the surface a velocity $2V$). In fact, detailed analysis shows that the tip has a velocity slightly greater than this simple estimate as a result of convergence effects (see below), but it is a good enough approximation for many cases.

Recent work has concentrated on how the collapse of neighbouring cavities is affected by their mutual interactions. This has raised much scientific interest as to possible focusing and

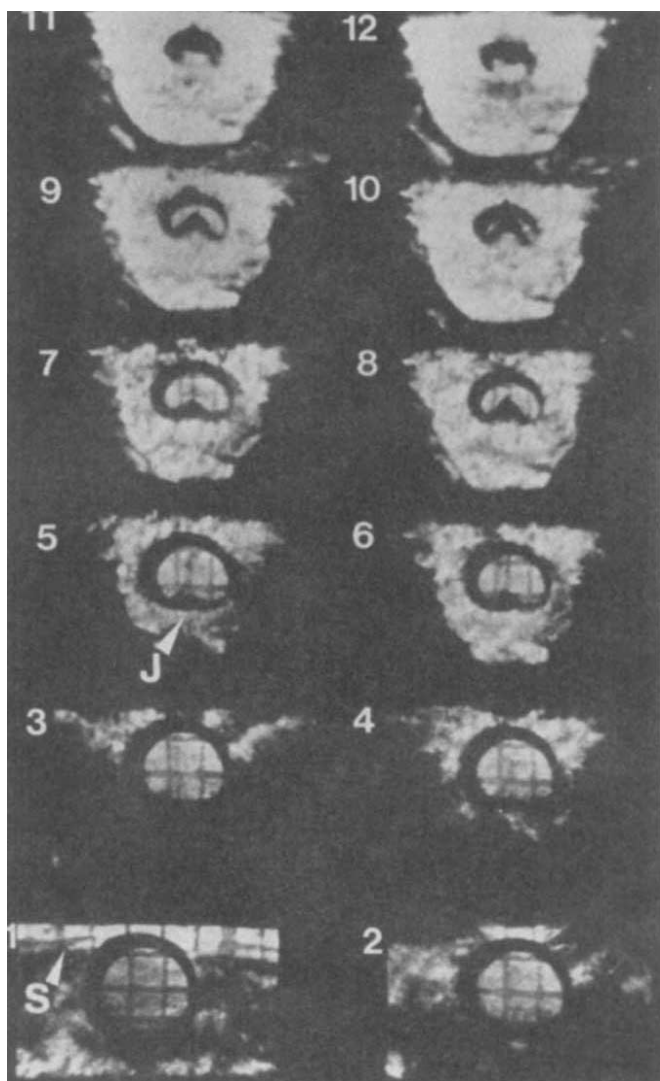


Fig. 2 Two-dimensional cavity of diameter 3 mm collapsed by a shock wave, labelled S. A jet, J, starts to form in frame 5. Inter-frame time, 0.96 μ s.

co-operative effects with clusters of cavities. For example, the pressure waves from cavities collapsing by the Rayleigh or Plesset and Chapman mechanism could pass over cavities and induce shock collapse and the much more violent jets that can form by this process^{11,12}. Some theoreticians^{13,14} have discussed a large hemisphere of cavities collapsing in consecutive shells from the outside inwards. At each stage, the energy of collapse is transferred to the inner shell, with a steady buildup of pressure resulting. They found that this increased the collapse energy of the cavities at the centre of the cluster by an order of magnitude. The present research gives information on this process.

A second area where cavity collapse is of interest is in the initiation and propagation of fast reaction in explosives (see, for example, refs 15 and 16). The rapid collapse can generate 'hot spots' by at least two mechanisms. The first is by adiabatic heating of the gas in the cavity. A detailed study of this was made¹⁷ using a water tank and bubbles of different sizes and gas contents collapsed adjacent to explosive crystals. Ignition of sensitive azide crystals was obtained by collapsing cavities of $>30 \mu\text{m}$ size by relatively weak shocks of 0.1 GPa strength. The cavities had to be collapsed in close, thermal contact with the explosive. The second ignition mechanism is from the 'hot spot' region produced by shock heating where the jet impinges⁸. This mechanism becomes more important at very high shock pressures (several GPa). It is clearly of interest to know the

position and size of the 'hot spots' for different shock collapse situations.

The fact that light may be emitted during certain violent cavitation processes has been known for over fifty years. This light emission is called sonoluminescence, in recognition of the fact that the appropriate cavitation conditions are usually achieved with a high-intensity sound field of amplitude sufficient to ensure that the cavitation is 'transient', rather than 'stable' (typically a few bars). For a review of sonoluminescence, see ref. 18. The light emission may also occur following cavitation of the liquid by a spark discharge, a laser pulse, or by blowing it through a Venturi tube. Sonoluminescence occurs in a variety of liquids, its intensity and spectrum depending on the nature of the solvent and the solute (including dissolved gases). Spectral measurements show that the luminescence originates mainly from the recombination of free radicals created within the high-temperature and high-pressure environment of a bubble undergoing an adiabatic compression, as may happen during transient cavitation. Recently there has been some concern as to the safety or otherwise of clinical diagnostic and therapeutic ultrasound (see, for example, ref. 19). Part of the concern arises from the possibility of free-radical creation within the patient undergoing examination or treatment. The fact that free radicals are indeed created within blood plasma and amniotic fluid by therapeutic generations (as evidenced by the emission of sonoluminescence) has recently been established^{20,21}.

Experimental

The two-dimensional gel technique²² has recently been improved and applied to a range of problems in fluid mechanics, such as impact with liquid drops²³, impact with liquid wedges^{24,25} and cavitation studies¹². It has been shown that at the high impact velocities used in these experiments the gel behaves like a liquid.

The gel layer is made by first dissolving 12% by weight of gelatine in water at 330 K and adding it to a vertical mould (dimensions $200 \times 200 \times 3 \text{ mm}$) also raised to this temperature. Before the addition of the heated gel, the mould faces are lightly greased and covered with a thin plastic film. After slow cooling, the mould is disassembled and the sheets placed horizontally. The layers, with plastic sheets attached, can be kept for several days. It is then relatively simple to cut out of the layer a chosen 'liquid' shape to be impacted, or to produce cavities by removing material, as in the present case. The part of the gel layer containing the cavity, or cavity array, is then placed between spaced glass blocks so that a striker, fired from a rectangular-bore gas gun, can be projected between them to impact the gel. Impact velocities of up to 300 m s^{-1} have been achieved to produce shock pressures of up to $\sim 0.6 \text{ GPa}$. The striker triggers an Imacon framing image converter camera by intersecting a laser beam just before impact. The shocks are visualized using Schlieren optics. A schematic diagram of the apparatus is given in Fig. 1.

In some experiments, the gel with its cavity array is viewed using a high-gain image intensifier system. The intensifier used (EMI type 9912) is a magnetically focused four-stage tube (of phosphor-photocathode sandwich construction) with a bi-alkali input photocathode and P11 phosphors. Run at 40 kV, the radiant power gain of the tube is sufficient to ensure that the majority of the single photons incident on the input photocathode (quantum efficiency 20%) which emit photoelectrons are recorded as distinct events by the camera. The gel was always photographed before impact so that the light output in the image intensifier record could be related to position in the cavity. Finally, some experiments were made using the high-speed camera to record the image on the output phosphor of the image intensifier.

Results

Figure 2 shows the collapse of a circular cavity of diameter $\sim 3 \text{ mm}$ by a shock wave of strength 0.26 GPa. This was produced

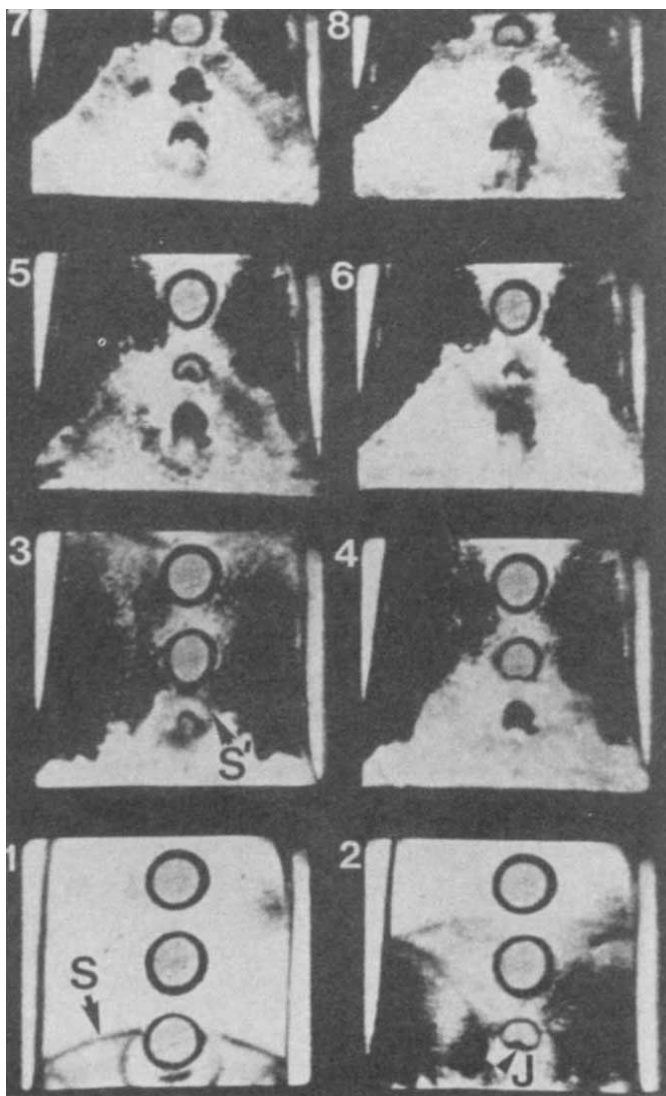


Fig. 3 Three cavities, of diameter 3 mm, perpendicular to a shock wave, *S*. *J* is the jet in the first cavity and *S'* the shock produced by the rebound of the first cavity. Note the step-by-step sequence of collapse. Inter-frame time, 5 μ s.

by slider impact, at a velocity of 150 m s⁻¹. The shock wave, labelled *S*, can be seen encircling the cavity in frame 1. A jet starts to form in frame 5 and its velocity averaged over the remaining frames is ~ 400 m s⁻¹.

Figure 3 is an example with three cavities in a vertical column. In frame 1, the shock *S* of strength 0.26 GPa can be seen interacting with the first cavity. By frame 2, the first cavity has started to collapse and a jet, labelled *J*, is forming. At this stage, the shock wave is close to the third cavity and reflected waves from the edge of the gel layer, now tensile, can be seen moving inwards, forming the darkened regions. The darkening is due either to minute cavities formed in the gel or detachment of the gel from the glass blocks. Note that the first cavity has shielded the second cavity, which is in the 'shadow' of the first, and that its collapse has not started. Careful examination of frame 2 shows that the central part of the primary shock is missing because of the shielding effect. Between frames 2 and 3, the first cavity reaches minimum volume and rebounds, forming the shock labelled *S'*. It is this shock which collapses the second cavity (frames 3–6). The rebound from this collapse then collapses the third cavity (frames 7 and 8). This sequence shows that a chain reaction along a line of cavities is thus feasible given the right conditions of shock strength, cavity diameter and

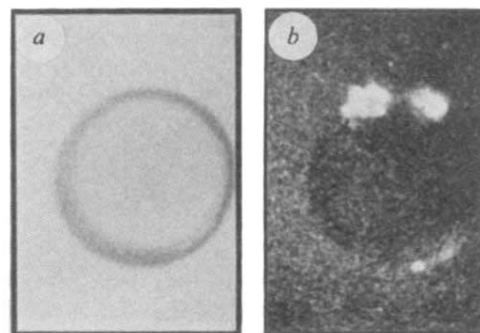


Fig. 4 *a*, A 3-mm-diameter cavity before collapse, and *b*, its time-integrated luminescence. The luminescence is concentrated in the lobes on either side of the jet.

spacing. For rectangular arrays of cavities, with the rows separated by no more than two or three cavity diameters, the collapse proceeds layer by layer¹². If the cavity spacing is much greater than this, then the initial shock diffracts around the first layer of cavities and the collapse processes approach those for isolated cavities.

Figure 4 shows a single 3-mm-diameter cavity collapsed under similar conditions to those in Fig. 2 but this time viewed through the image intensifier. Figure 4*a* shows the cavity before collapse in ambient light, and Fig. 4*b* the time-integrated luminescence from the collapsed cavity. A small amount of background lighting has been added to reveal the shape of the uncollapsed cavity. There is a bright spot on the side of the cavity where the shock first interacts, but this did not appear in all sequences and is probably an artefact caused by an imperfection in the gel surface. But it is clear that most light is produced from the two lobe-shaped regions of trapped gas on either side of the jet. These conform to the shape of the cavity wall during the late stages of collapse (see frame 10 of Fig. 2). Note that there are several bright sources within each of the lobe-shaped regions. These observations were confirmed in other experiments, including ones in which a high-speed camera was used to record the changes on the screen of the image intensifier. A similar configuration to that of Fig. 3, with three cavities in line, was also studied. The brightest luminescence continued to be concentrated in the lobes, but there was a small decrease in the intensity along the chain (see Conclusions). In none of the experiments was there any indication that a 'hot spot' formed in the region where the jet struck the cavity wall.

Figure 5*a* and *b* illustrate the significance of the above experiments to explosive initiation with selected frames taken from high-speed sequences. In Fig. 5*a*, a cavity was added to the gel layer as before, but this time a small amount (~ 10 mg) of the primary explosive, silver nitrotriazole, was positioned in the far cavity wall. The shock of strength 0.25 GPa passes upwards and the jet, labelled *J*, starts to form in frame 2. Ignition, labelled *I*, occurs in frame 3 with the ignition site to the left of where the jet impacts (that is, near one of the lobes). Figure 5*b* shows the effect of having jet impact alone. This was achieved by having a semi-circular cavity that allows a jet to form (frame 2 onwards), but the gas is vented. In this case, there was no evidence of ignition either during the time of the sequence (not all shown here), or by examining the glass blocks afterwards.

Conclusions

The problem of the shock-wave collapse of a cavity has been considered numerically⁸ and analytically¹⁰. The latter approach is briefly noted here because it helps in understanding the mechanics of the jet production. Figure 6 shows a planar shock wave interacting with a circular cavity. Reflection of the shock wave, *S*, occurs at the free liquid surface to produce a corner wave, *C*, and a reflected wave *R*. It can be shown that the velocity imparted to the free surface is given by $V_r = 2V \sin \theta$, where *V*

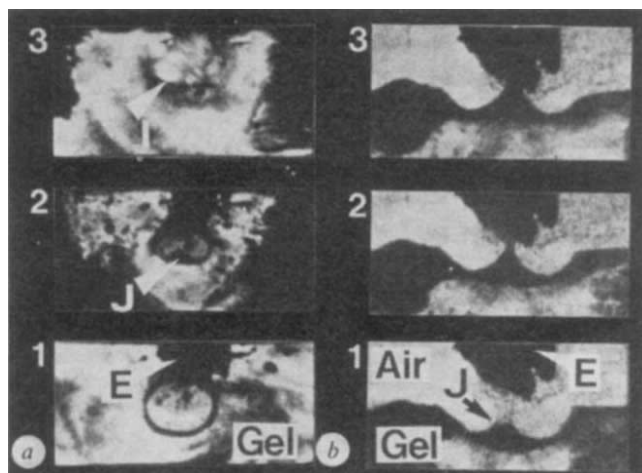


Fig. 5 *a*, The collapse of a 3-mm cavity onto an explosive, labelled E. The jet, J, forms in frame 2. Ignition, I, occurs near one of the lobes of compressed gas. Inter-frame time, 5 μ s. *b*, A semi-circular cavity is used and frame 1 shows the gel/air/explosive configuration. A jet of velocity ~ 400 m s $^{-1}$ forms, labelled J, and strikes the explosive but the gas is vented. Ignition of the explosive does not take place. Inter-frame time, 5 μ s.

is the particle velocity behind the shock. Linearized shock relations are assumed in the liquid of the form $P = \rho CV$. To calculate the perturbed surface shape, Lesser¹⁰ then assumes that the particles in the liquid surface travel with the initial velocity imparted to them. The resulting shapes for various non-dimensional times and for $V = 150$ m s $^{-1}$ ($C = 1,500$ m s $^{-1}$) are given in Fig. 6*b*, the non-dimensional time $\bar{t}$ being defined as $\bar{t} = tC/R$, where R is the cavity radius. The maximum velocity of the cavity wall is attained at the centre of the cavity, where the wall velocity is twice the particle velocity behind the shock. After a time $\bar{t} = 5$, corresponding to the time when particles of liquid from opposite sides of the cavity meet, the cavity wall forms a cusp. The shapes that the Lesser¹⁰ treatment predict are reasonably close to those we observe, but his jet tip of velocity $2V$ is likely to be low as he does not allow for convergence and nonlinear effects: the numerical treatment⁸ for an 8.5-GPa shock passing over a cavity in nitromethane suggests that convergence effects can increase the final jet velocity by a factor of 1.5 times the free surface velocity $2V$. In our experiments at 0.26 GPa, $2V = 300$ m s $^{-1}$, but the jet velocities were measured at ~ 400 m s $^{-1}$. This is reasonable as the convergence effects are likely to be less than a factor 1.5 at the lower pressures. Jet velocities of this magnitude would clearly have damage potential. For example, if the cavity collapse took place near a metal surface, pressures of the order of 1 GPa would be generated.

The experiments with multiple cavities show that the collapse is a step-by-step process, with the velocity of the chain reaction related to the times of collapse of individual cavities (or layers of cavities in a two- or three-dimensional array) and their separation, and only indirectly to the shock velocity. In linear

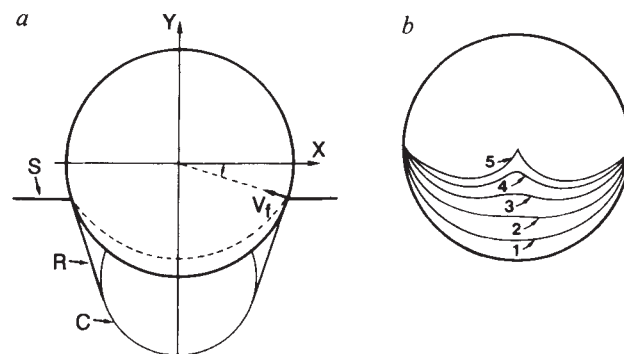


Fig. 6 *a*, Shock-wave interaction with a circular cavity. *b*, Resulting shapes of cavity wall for non-dimensional times $\bar{t} = 1$ to 5. For explanation of notation, see text.

or rectangular arrays a high fraction of the collapse energy (~ 80 – 90% , estimated from the ratio of the jet velocities squared) is transmitted to the next layer. That this is a reasonable estimate is confirmed by luminescence sequences (not illustrated), which show that the luminescence only decays by a small amount from one cavity to the next. With a hemispherical cavity cluster¹³, an amplification towards the centre would be possible.

The luminescence strongly suggests that the gas reaches high temperatures in the two lobes. Measurements of the volume of gas luminescing show that luminescence takes place when the volume has been compressed to less than 10% of its initial volume. Assuming that this was performed adiabatically, this predicts a temperature in excess of 750 K. This is consistent with a gas luminescence arising from free-radical creation and radiative recombination²⁶. It is also consistent with ignition of the primary explosive (Fig. 5*a*). The luminescence shows a definite structure within some brighter regions, however. A model in which the shock waves in the gas are analysed could be more appropriate¹⁰.

At the shock pressures used in the present study (up to ~ 0.6 GPa), the 'hot spots' formed in the gas; there was no evidence of a 'hot spot' in the liquid where the jet impacted. Mader's numerical work⁸ suggests that the jet 'hot spot' mechanism dominates when pressures of several GPa are achieved. In future experiments we plan to explore the higher-pressure region. The experiments with the primary explosive (Fig. 5) have confirmed the importance of the gas compression at shock pressures in the range studied.

The authors are grateful to a referee for pointing out the interesting similarities between the pictures illustrated here and those showing the interaction of a weak shock with the cylindrical gas inhomogeneity embedded in a gaseous medium²⁷. For the case of helium embedded in air, the shock interaction caused involution of the interface and jet formation in a similar manner to that illustrated in Fig. 2. The instability was treated as an example of the shock-induced Rayleigh–Taylor or Richtmeyer–Meshkov instability.

J.P.D. thanks SERC and Girton College, Cambridge for research fellowships.

Received 9 November 1987; accepted 11 February 1988.

1. Rayleigh, Lord *Phil. Mag.* **34**, 94–98 (1917).
2. Plesset, M. S. & Prosperetti, A. A. *Rev. Fluid Mech.* **9**, 145–183 (1971).
3. Mörch, K. A. in *Erosion* (ed. Preece, C. M.) 309–353 (Academic, New York, 1979).
4. Hickling, R. & Plesset, M. S. *Physics Fluids* **7**, 7–14 (1964).
5. Kornfeld, M. & Suvorov, L. J. *Appl. Phys.* **15**, 495–506 (1944).
6. Plesset, M. S. & Chapman, R. B. J. *Fluid Mech.* **47**, 283–290 (1971).
7. Lauterborn, W. & Bolle, H. J. *Fluid Mech.* **72**, 391–400 (1975).
8. Mader, C. L. *Physics Fluids* **8**, 1811–1816 (1965).
9. Mader, C. L. *Numerical Modelling of Detonations* (Univ. of California Press, Berkeley, 1979).
10. Lesser, M. B. & Ezekoye, O. A. *Proc. 7th Int. Conf. on Erosion by Liquid and Solid Impact* (eds Field, J. E. & Dear, J. P.) (Cavendish Laboratory, Cambridge, UK, 1987).
11. Tulin, M. P. in *L.I. Sedov's 60th Anniversary Vol: Problems of Hydrodynamics and Continuum Mechanics* (Society for Industrial and Applied Mathematics, Philadelphia, 725–747 1969).
12. Dear, J. P. & Field, J. E. J. *Fluid Mech.* (in the press).
13. Hansson, I. & Mörch, K. A. J. *Appl. Phys.* **51**, 4651–4658 (1980).
14. Hansson, I., Kendrinskii, V. & Mörch, K. A. J. *Phys. D* **15**, 1725–1734 (1982).

15. Bowden, F. P. & Yoffe, A. D. *Initiation and Growth of Explosions in Liquids and Solids* (Cambridge University Press, 1985).
16. Bowden, F. P. & Yoffe, A. D. *Fast Reactions in Solids* (Butterworths, London, 1958).
17. Chaudhri, M. M. & Field, J. E. *Proc. R. Soc. A* **340**, 13–128 (1974).
18. Walton, A. J. & Reynolds, G. T. *Adv. Phys.* **33**, 595–660 (1984).
19. *Biological Effects of Ultrasound: Mechanism and Clinical Implications* Report no. 74 (National Council on Radiation Protection, Bethesda, 1983).
20. Crum, L. A., Daniels, S., Dyson, M., ter Haar, G. R. & Walton, A. J. *Proc. Instn Acoust.* **8**, 137–146 (1986).
21. Crum, L. A. et al. *J. Ultrasound Med.* **6**, 643–647 (1987).
22. Field, J. E., Lesser, M. B. & Davies, P. N. H. *Proc. 5th Int. Conf. on Erosion by Liquid and Solid Impact* (ed. Field, J. E.) (Cavendish Laboratory, Cambridge, UK, 1979).
23. Field, J. E., Dear, J. P. & Ögren, J.-E. J. *Appl. Phys.* (in the press).
24. Field, J. E., Dear, J. P., Davies, P. N. H. & Finnström, M. *Proc. 6th Int. Conf. on Erosion by Liquid and Solid Impact* (eds Field, J. E. & Corney, N. S.) (Cavendish Laboratory, Cambridge, UK, 1983).
25. Field, J. E., Lesser, M. B. & Dear, J. P. *Proc. R. Soc. A* **401**, 225–249 (1985).
26. Sehgal, C., Sutherland, R. G. & Verall, R. E. J. *Phys. Chem.* **84**, 388–390 (1980).
27. Haas, J.-F. & Sturtevant, B. J. *Fluid Mech.* **181**, 41–76 (1987).

A specific inhibitor of the *ran1*⁺ protein kinase regulates entry into meiosis in *Schizosaccharomyces pombe*

Maureen McLeod & David Beach

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

In fission yeast, meiosis is initiated by transcriptional activation of the mei3⁺ gene, under the combined influence of the four mating-type genes. The product of the mei3⁺ gene acts as a critical meiotic inducer by binding non-covalently to a newly identified protein kinase encoded by the ran1⁺ gene and inhibiting its enzymatic activity. Inactivation of the ran1⁺ protein kinase is both necessary and sufficient to divert a vegetative cell from mitotic division to meiotic differentiation.

SEXUAL reproduction is almost universal among eukaryotes, but the molecular basis of the developmental switch from somatic mitotic cell division to meiosis and gametogenesis is not well understood. The unicellular ascomycete, *Schizosaccharomyces pombe*, is particularly suitable for the study of this process at the molecular level due to the ease with which it can be manipulated with the techniques of both classical genetics and recombinant DNA^{1,2}. Genetic analysis has identified both activators and inhibitors of meiosis. These include the *mat* genes¹, *mei3*⁺ (ref. 3) and *ran1*⁺ (refs. 4, 5), each of which has been physically isolated⁶⁻¹⁰. During meiosis, the mating-type genes act to induce transcription of the *mei3*⁺ gene⁹. This gene is required for meiosis³ and its expression alone is sufficient to initiate meiosis in a vegetative cell⁹. Here we show that the *mei3*⁺ gene product is an inhibitor of a newly identified protein kinase encoded by the *ran1*⁺ gene¹¹. This protein kinase is a critical negative regulator of meiosis and its inactivation causes immediate stimulation of meiotic development⁴⁻⁶. Thus, we establish that the *mei3*⁺ gene product acts in meiosis as the

biochemical link that brings the *ran1*⁺-encoded protein kinase under control of the mating-type genes.

Meiotic control

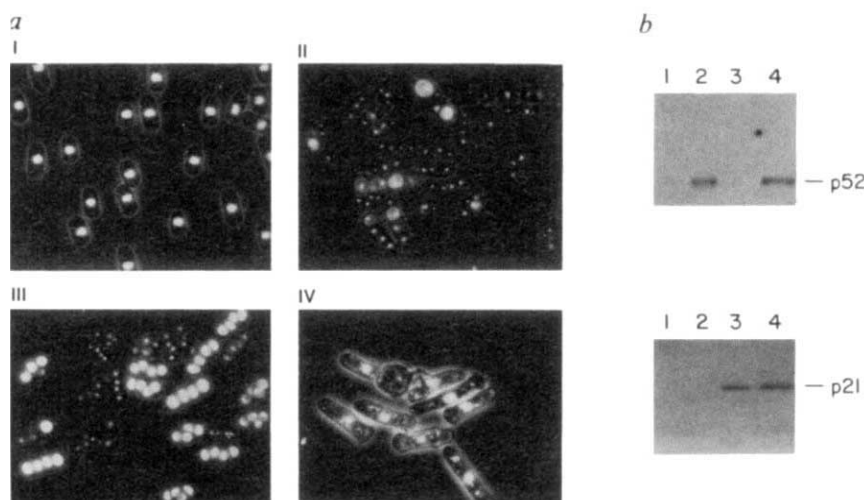
In *Schiz. pombe*, sexual differentiation is tightly regulated by both cell-type and nutritional signals. Haploid cells propagate vegetatively as one of two mating types termed *h*⁺ or *h*⁻ (ref. 1). Under conditions of nutrient deprivation, cells of opposite mating type undergo sexual fusion giving rise to a temporary *h*⁺/*h*⁻ diploid zygote¹². The *h*⁺/*h*⁻ zygote usually enters meiosis directly and develops into an ascus containing four haploid spores, but under certain conditions meiotically competent diploids may be induced to propagate vegetatively before sporulating¹³ (Fig. 1a).

Mating type (*h*⁺ or *h*⁻) is determined by the allele expressed at the *mat1* locus of the mating-type region, respectively either *mat1-P* or *mat1-M* (refs 8, 14). Neither *mat1-P* nor *mat1-M* is a single gene; each contains two divergently transcribed genes known as *Pi* and *Pc*, and *Mi* and *Mc*¹⁵. *Pc* and *Mc* are sufficient

Fig. 1 Meiotic mutants of *Schiz. pombe*. **a**, Panel I, *h*⁻ haploid strain cultured in complete medium at 33 °C (972); Panel II, *h*⁻ *ran1*.114 mutant undergoing haploid meiosis and sporulation in complete medium at 33 °C (SP399); Panel III, *h*⁺/*h*⁻ diploid strain undergoing meiosis in nitrogen limited medium (SP272); Panel IV, Diploid *ran1*⁺O.P. strain cultured under conditions of nitrogen starvation (SP632). Meiosis is inhibited at least 1,000-fold in this strain. **b**, Immunoblots of whole cell lysates of wild-type (972, lane 1); *ran1*⁺O.P. (SP618, lane 2); *mei3*⁺O.P.*mei2* (SP799, lane 3) and *ran1*⁺O.P.*mei3*⁺O.P.*mei2* (SP801, lane 4) strains probed with either monoclonal antibody R30 to p52^{ran1} (top) or antibody N8 to p21^{mei3} (bottom). Note that only at higher levels of exposure is p52^{ran1} detectable in the non-overproducer strains.

Methods. Strain constructions are described in Table 1. Cells cultured in standard minimal medium⁴¹ or meiotically induced⁶ were fixed in ethanol and stained with propidium iodide for photomicroscopy⁹.

Extracts for Western blots were prepared by vortexing pelleted cells in buffer 1 (50 mM Tris pH 8.0, 5 mM EDTA, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulphonylfluoride (PMSF)) in the presence of glass beads (0.45 mm; Braun) for 3 min at 4 °C. Lysates were diluted in buffer 2 (Buffer 1 plus 0.1% SDS, 0.5% deoxycholate, 1.0% Triton-X 100) and centrifuged at 100,000g for 30 min. Supernatant containing 100 µg protein was subjected to 7.5–15.0% SDS-PAGE⁴² and analysed by Western blotting⁹. Under these conditions p52^{ran1} is undetectable in wild-type cells. Monoclonal antibody N8 has been described⁹. Antibody R30 was obtained by immunizing mice with p52^{ran1} purified from a bacterial strain producing high amounts of the protein¹¹. Briefly, p52^{ran1} obtained as the insoluble fraction of whole bacterial lysate, was resolubilized in TU buffer (50 mM Tris pH 8.0, 6 M urea) and applied to a DEAE cellulose (DE-52, Whatman) column that had been equilibrated in TU buffer. Protein was eluted in a gradient of 0–1.0 M NaCl in TU buffer and those fractions containing p52^{ran1} were visualized by SDS-PAGE, pooled and dialysed against 50 mM Tris pH 8.0, 2 mM EDTA. BALB/c female mice were immunized with this material and hybridomas were produced as described⁹.



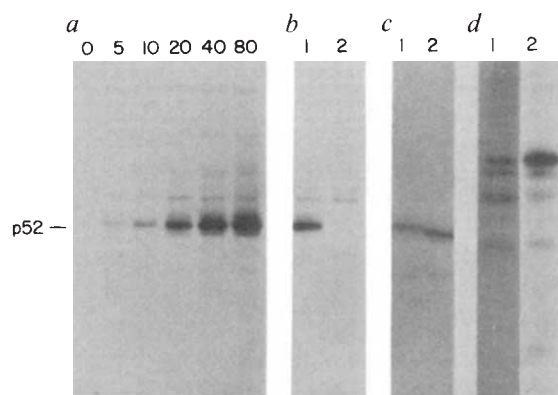


Fig. 2 Autophosphorylation of p52^{ran1}. *a*, Kinase assay of p52^{ran1} immunoprecipitated from a *ran1*⁺O.P. strain (SP618) with anti-p52 polyclonal antibody, incubated with ³²P-ATP for the indicated times and analysed by SDS-PAGE. Numbers, time in min. *b*, Kinase assay of p52^{ran1} immunoprecipitated from *ran1*⁺O.P. (SP618, lane 1) or *ran1*^{Arg47}O.P. (SP802, lane 2) strains. *c*, Western immunoblot of anti-p52 immunoprecipitates of lysate prepared from *ran1*⁺O.P. (lane 1) or *ran1*^{Arg47}O.P. (lane 2) strains, probed with anti-p52 MAb R30. *d*, Silver stain of *N*-chlorosuccinimide/urea cleavage products of p52^{ran1} purified from *E. coli* (lane 1) and autoradiograph of *N*-chlorosuccinimide/urea cleavage products of ³²P-p52 phosphorylated in an *in vitro* kinase reaction. Note that these digestions are partial and that the uppermost band is undegraded p52^{ran1} (lane 2).

Methods. Strain SP802 (*h90leu1.32ura4ade6.210mei2ran1.4:ran1*^{Arg47}O.P.) was constructed as follows. An Arg for Lys substitution was introduced into the cloned *ran1* gene (plasmid pRAN1.39) at residue 47 by site-directed oligonucleotide mutagenesis²² using the following mutagenic primer: 5'TGTATTTTCGGGAATGACGTAT3', and this was used for homologous gene recombination in strain SP236 (*h90leu1.32ade6.210ran1.4mei2f.14*) by transformation to stable Leu⁺ to give strain SP802. Kinase assays were performed with polyclonal anti-p52 precipitates from cell lysates. 1.0 mg cell lysate (Fig. 1) was incubated at 4°C for 1 h with 10 µl anti-p52 antiserum (see below). Following precipitation of immune complexes using 20 µl protein A-sepharose (Pharmacia) the immuno-beads were washed twice in buffer 2 (Fig. 1), once in 50 mM Tris pH 8.0 mM EDTA and finally in KAB buffer (50 mM Tris pH 8.0, 1 mM EDTA, 5 mM NaCl₂, 1 mM DTT). The kinase reactions were started by the addition of ATP to 40 µM and of 10 µCi ³²P-ATP in a final volume of 40 µl KAB buffer. After 30 min (unless otherwise indicated) at 30°C, reactions were terminated by addition of 2× Laemmli sample buffer⁴² containing 20 mM EDTA. Proteins were separated on a 7.5%–15% SDS-polyacrylamide gel and visualized by exposure to photographic film for 8 h with an intensifying screen. Proteolytic cleavage of proteins was by exposure to *N*-chlorosuccinimide/urea^{25,35}. Anti-p52 antibody was prepared using p52^{ran1} purified from bacteria (Fig. 1) as antigen. Rabbits were immunized using 100 µg p52^{ran1} in complete Freund's adjuvant. Subsequent injections contained 100 µg protein in incomplete Freund's adjuvant and were administered until significant anti-p52^{ran1} immunoreactivity was detected by immunoprecipitation of ³⁵S-methionine labelled protein from *ran1*⁺O.P. (SP618) lysate. Serum was prepared by centrifugation of rabbit blood at 3,000 r.p.m. for 15 min.

to confer an *h*⁺ or *h*[−] mating phenotype respectively, but all four genes are required for meiotic competence in an *h*⁺/*h*[−] diploid strain¹⁵. Each of the four *mat* genes is transcriptionally regulated by nutritional signals, full activation being achieved only under conditions of nitrogen starvation¹⁵. Consequently, no expression of cell-type differentiation is observed in either haploid or diploid strains, except under nutritional stress.

In common with the *MAT* genes of *Saccharomyces cerevisiae*^{16,17}, the mating-type genes of *Schiz. pombe* act as transcriptional regulators of at least one, and probably many, unlinked genes. *mei3*⁺ is controlled by the *mat* locus and plays

a critical role in meiosis^{3,9}. A strain carrying a null-allele of *mei3* is viable but meiotically defective⁹ (Table 1). Following sexual conjugation, *mei3* zygotes become arrested very early in meiotic development, before the initiation of premeiotic DNA replication¹⁸. The *mei3*⁺ gene encodes a transcript of 1,300 nucleotides and a translational product of relative molecular mass (*M*_r) 21,000 (21K) both of which are detectable only during meiosis⁹. Point mutations within *mat* that block meiosis prevent transcription of *mei3*⁺ (ref. 9). At present it is not known whether the *mat* gene products directly regulate *mei3*⁺ transcription or act through an intermediate.

In non-meiotic cells *mei3*⁺ is transcriptionally silent, but vegetative expression can be induced by fusing the gene to a heterologous promoter and introducing the plasmid construct into fission yeast. Expression of *mei3*⁺ in either *h*⁺ or *h*[−] vegetative haploid cells has dramatic consequences. Cellular growth and mitotic division cease abruptly and haploid meiosis and sporulation follow⁹ (Table 1). Because expression of the *mei3*⁺ gene alone is sufficient to induce meiosis and sporulation in a vegetative cell, transcriptional activation of *mei3*⁺ may be the main meiotic role of the four *mat* genes.

Meiosis in *Schiz. pombe* is regulated not only by meiotic activators but also by a meiotic inhibitor. The *ran1*⁺ gene functions as a critical negative regulator of sexual conjugation, meiosis and sporulation^{4,6}. Only its role in meiosis and sporulation is considered here. A *ran1ts* mutant is a recessive temperature-sensitive lethal that ceases vegetative growth and sporulates under restrictive conditions^{4,6} (Fig. 1*a*, Table 1). This behaviour is the same as that of a haploid strain that has been induced to express *mei3*⁺ (ref. 9) and mimics that of an *h*⁺/*h*[−] diploid entering a normal meiosis^{6,19}. However, the viability of spores resulting from haploid meiosis in a *ran1ts* mutant is very low, presumably due to abnormalities of chromosome segregation.

These observations are consistent with two quite different models. Either *ran1*⁺ might act as a meiotic inhibitor by repressing *mei3*⁺, or equally *mei3*⁺ might serve to inhibit the *ran1*⁺ gene or its protein product. A simple epistasis test reveals the relationship between these two genes. The *ran1tsmei3* double mutant displays the meiotic phenotype of the *ran1ts* rather than the *mei3* mutant^{6,19} (Table 1). Therefore *ran1ts* is epistatic to *mei3*, suggesting that *ran1*⁺ acts after *mei3*⁺ in the meiotic pathway, and that the role of *mei3*⁺ might be to inhibit *ran1*⁺ activity (refs 6, 9, 19). However, neither the abundance of the *ran1*⁺ transcript¹¹ nor its protein product decline during meiosis under the influence of *mei3*⁺ (M.M., unpublished). We show below that the interaction between these two meiotic regulators is at the protein–protein level.

Gene interaction

Because haploid meiosis occurs in *ran1ts* mutants it has been assumed that inhibition of *ran1*⁺ activity is a necessary step in normal meiosis^{4,6}. We have obtained two new pieces of genetic evidence that support this conclusion and also demonstrate that *mei3*⁺ is a component of the inhibitory mechanism.

First, if inactivation of *ran1*⁺ is essential for meiotic initiation, overexpression of the gene might be expected to block meiosis by overwhelming the capacity of the *ran1*⁺-inhibiting pathway. To test this hypothesis, *ran1*⁺ was overexpressed approximately 30-fold by fusing the coding sequence of the gene to the stronger promoter of the *Schiz. pombe* alcohol dehydrogenase gene²⁰ and introducing this construction into wild-type *Schiz. pombe* (the overproducer alleles of this and other genes are referred to hereafter as O.P.). The identity of the *ran1*⁺O.P. strain was confirmed by Western blotting²¹ using a monoclonal antibody prepared against the *ran1*⁺ gene product (Fig. 1*b*). *Schiz. pombe* cells carrying *ran1*⁺O.P. were totally sterile as haploids (SP618) and *h*⁺/*h*[−] *ran1*⁺O.P. diploids (SP632), constructed by protoplast fusion *in vitro*, were profoundly meiotically defective (Fig. 1*a*, Table 1).

Second, if overexpression of *ran1*⁺ is indeed inhibitory to meiosis because the capacity of a *ran1*⁺-inactivating pathway is exceeded, and if the *mei3*⁺ gene product is the limiting component of this pathway, overexpression of the *mei3*⁺ gene is predicted to counteract the meiotic defect of *ran1*⁺O.P., even to the point of provoking haploid meiosis. This was precisely the result obtained. A *ran1*⁺O.P.*mei3*⁺O.P. strain could not be propagated vegetatively, because it entered haploid meiosis and sporulation non-conditionally (Table 1). The double-overexpressor strain could, however, be obtained in a *mei2* mutant background. This suppressor mutation has been shown to inhibit normal meiosis³ and also to suppress expression of both the *ran1ts* and *mei3*⁺O.P. haploid meiotic phenotypes, allowing strains carrying either allele to propagate vegetatively^{6,9,19}.

The *mei3*⁺O.P. mutation is epistatic to *ran1*⁺O.P. Thus, the double-overproducer cell is phenotypically identical to both a *mei3*⁺O.P. strain⁹ and to a *ran1* null-mutant⁶, despite the abundance of the *ran1*⁺ gene product in the double-overproducer *mei2* strain (SP801, Fig. 1b). We show below that this protein is biochemically inactive in the presence of *mei3*⁺ gene product.

The *ran1*⁺ protein kinase

We have previously determined the nucleotide sequence of the *ran1*⁺ gene and have noted significant homologies between its

Table 1 Summary of meiotic mutants

Genotype	Sporulation in <i>h</i> ⁺ / <i>h</i> ⁻ diploid	Sporulation in haploid**
<i>ran1</i> ⁺ <i>mei3</i> ⁺	+	—
<i>ran1</i> ⁺ <i>mei3</i> [*]	—	—
<i>ran1mei3</i> ⁺⁺	n.a. [#]	+
<i>ran1mei3</i> [‡]	n.a. [#]	+
<i>ran1</i> ⁺ O.P. <i>mei3</i> [§]	—	—
<i>ran1</i> ⁺ <i>mei3</i> ⁺ O.P.	n.a. [#]	+
<i>ran1</i> ⁺ O.P. <i>mei3</i> ⁺ O.P. [¶]	n.a. [#]	+

Plasmid pRAN1.39 was used to obtain overexpression of p52^{ran1} in fission yeast. This plasmid carries the *ran1*⁺ coding sequence¹¹ under control of the constitutive *adh* promoter in the integrating expression vector pART5. This is identical to the previously described pART3⁹, except that the *ars* element is missing. The *ran1*⁺O.P. strain (SP618, *h*⁹⁰*leu1.32ade6.216ran1*⁺::pRAN1.39) was obtained by introduction of pRAN1.39 into SP66 (*h*⁹⁰*leu1.32ade6.216*) by transformation³⁷ and selection of stable leucine prototrophs. A diploid strain carrying *ran1*⁺O.P. (SP632, *h*⁹⁰*leu1.32ade6.210/h*⁹⁰*leu1.32ade6.216ran1*⁺::pRAN1.39) was constructed by protoplast fusion³⁸ of SP618 with SP67 (*h*⁹⁰*leu1.32ade6.210*) followed by selection of adenine prototrophs³⁹. High level expression of p21^{mei3} was obtained by using pMEI3.20, which contains the *mei3*⁺ gene under control of the *adh* promoter in the vector pART4. This is identical to pART5, except that it carries the *ura3* gene of *Sacc. cerevisiae* that rescues *ura4* mutants of *Schiz. pombe*⁴⁰. It has been shown previously that high level expression of p21^{mei3} is lethal to wild-type vegetative cells because it induces haploid meiosis and sporulation⁹. The lethality of the *mei3*⁺O.P. allele is fully suppressed in a *mei2* mutant background⁹. Likewise, introduction of pMEI3.20 into cells carrying *ran1*⁺O.P. was lethal. Although in this instance haploid meiosis was not directly observed it is inferred that it occurs because the plasmid could readily be introduced into a *ran1*⁺O.P.*mei2* mutant strain with no adverse effect. The two strains carrying *mei3*⁺O.P. used in this study were SP799 (*h*⁹⁰*leu1.32ura4ade6.210mei2mei3*⁺::pMEI3.20) and SP801 (*h*⁹⁰*leu1.32ura4ade6.210mei2ran1*⁺::pRAN1.39*mei3*⁺::pMEI3.20).

* From ref. 3.

† From refs 4–6.

‡ From refs 6, 19.

§ O.P. Overproduction of indicated gene under control of the *adh* promoter (see text).

|| From ref. 9.

¶ The phenotype of this strain is inferred, see above.

Not applicable, because a strain undergoing haploid meiosis is not assayed as a diploid.

** Haploid may be either *h*⁺ or *h*⁻.

52K product and the family of protein kinases¹¹. To test whether p52^{ran1} is also a protein kinase, the protein was purified from a *ran1*⁺-expressing strain of *Escherichia coli*¹¹ and used to prepare polyclonal antibodies (Fig. 2). p52^{ran1}-dependent kinase activity was not detectable in anti-p52 immunoprecipitates of lysates prepared from wild-type *Schiz. pombe*, due to the low abundance of the protein in fission yeast. However, phosphorylation of a 52K protein was observed in immunoprecipitates of a whole cell lysate from a *ran1*⁺O.P. strain (SP618, Fig. 2a).

To confirm that phosphorylation of the 52K protein is dependent on the enzymatic activity of p52^{ran1}, an arginine for lysine substitution was introduced by site-directed mutagenesis²² of the *ran1*⁺ gene at the predicted site of ATP binding (Lys 47)¹¹. The equivalent mutation in various other protein kinases abolishes their biochemical activity^{23,24}. The mutant protein was overexpressed in fission yeast. No phosphorylation of a 52K band was detected in immunoprecipitates prepared from this strain (SP802, Fig. 2b), even though p52^{ran1Arg 47} was expressed at high levels and was recognized by the polyclonal anti-p52 serum to the same extent as the wild-type protein (Fig. 2c).

The previous experiment directly demonstrates that phosphorylation of a 52K protein is dependent on the activity of *ran1*⁺ and suggests that it might be autophosphorylation of p52^{ran1} itself. This was formally demonstrated by excising the 52K ³²P-labelled band from a gel and exposing it to *N*-chlorosuccinimide, a reagent that specifically cleaves polypeptides at tryptophan residues²⁵. We expected p52^{ran1} to have five such residues¹¹. The pattern of partial cleavage fragments of the radiolabelled protein was identical to that of unlabelled p52^{ran1} purified from *E. coli* and visualized by silver stain (Fig. 2d). This indicates that the two proteins are the same and, as each of the peptide fragments contain phosphoaminoacids, that p52^{ran1} is probably autophosphorylated at more than one site.

Kinase inhibition

Although *ran1*⁺O.P.*mei3*⁺O.P.*mei2* cells express p52^{ran1} at a high level (Fig. 1b), without the *mei2* suppressor mutation haploid sporulation would occur (Table 1). This phenotype is like that of cells that lack *ran1*⁺, rather than those that overexpress the protein. This observation prompted us to assay the kinase activity of p52^{ran1} in this strain. The autophosphorylation activity of p52^{ran1} was compared in immunoprecipitates of lysates prepared from *ran1*⁺O.P.*mei2* (SP800), *ran1*^{Arg 47}O.P.*mei2* (SP802) and *ran1*⁺O.P.*mei3*⁺O.P.*mei2* (SP801) strains. As described above, p52^{ran1} autophosphorylation was undetectable in the *ran1*^{Arg 47} mutant. More interest-

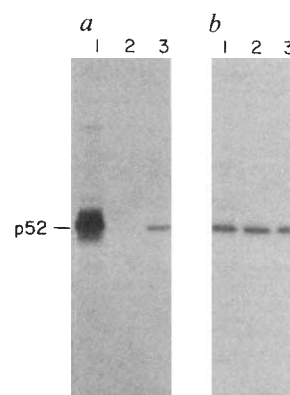


Fig. 3 Inhibition of p52^{ran1} autophosphorylation. Lysates prepared from *ran1*⁺O.P.*mei2* (SP800, lane 1), *ran1*^{Arg 47}O.P.*mei2* (SP802, lane 2) or *ran1*⁺O.P.*mei3*⁺O.P.*mei2* (SP801, lane 3) strains were immunoprecipitated with polyclonal anti-p52 antibody. Immunoprecipitated material was assayed for kinase activity (a) or for the presence of p52^{ran1} detected by immunoblotting with antibody R30 (b).

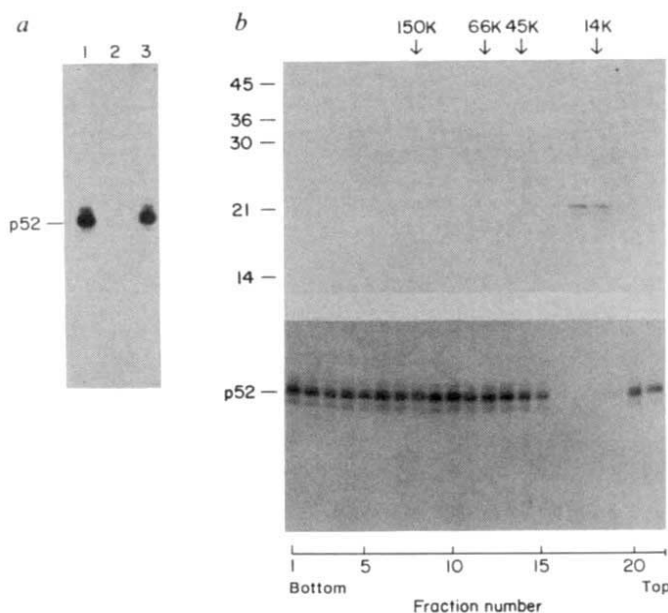


Fig. 4 Inhibition of $p52^{\text{ran1}}$ kinase activity by $p21^{\text{mei3}}$. **a**, The $\text{ran1}^+ \text{O.P.}$ (SP618) lysate was immunoprecipitated with polyclonal anti-p52 antibody. Before the addition of ATP to the kinase reaction, either no material (lane 1), 0.5 μg bacterial lysate containing $p21^{\text{mei3}}$ (lane 2) or 0.5 μg bacterial lysate without $p21^{\text{mei3}}$ (lane 3) was added to the immunocomplexes. **b**, Sedimentation of purified $p21^{\text{mei3}}$ analysed by SDS-PAGE (top) and $p52^{\text{ran1}}$ kinase inhibitory activity (bottom) thorough a glycerol gradient. Numbers marked by arrows, position of molecular weight markers in the gradient. **Methods.** Soluble extract was prepared from *E. coli* strain BL21(DE3)LysS or from the same strain carrying the plasmid pMEI3.18 (ref. 9). This plasmid contains the mei3^+ coding sequence under the control of an inducible promoter of the bacteriophage T7⁹. To purify $p21^{\text{mei3}}$, 70 mg soluble extract from the $p21^{\text{mei3}}$ expressing strain was loaded onto a hydroxylapatite (Bio-Rad) column equilibrated in 50 mM Tris pH 7.4, 100 mM NaCl, 10 mM KH_2PO_4 . The flow through fraction was dialysed against buffer B (50 mM Tris pH 8.0, 1 mM EDTA) and applied to a DEAE cellulose column (DE-52, Whatman). Protein was eluted with a gradient of 0–1.0 M NaCl in buffer B. Fractions enriched for $p21^{\text{mei3}}$ were visualized by Coomassie blue staining of SDS-PAGE and were applied to a hydroxylapatite column pre-equilibrated in buffer C (50 mM Tris pH 8.0, 150 mM NaCl). Protein was eluted with buffer C containing 10 mM KH_2PO_4 . Following dialysis against buffer B, this material was applied to a monoQ HR5/5 column (Pharmacia) equilibrated with buffer B and protein was eluted in a gradient of 0–1.0 M NaCl in buffer B. Each fraction was subjected to SDS-PAGE, visualized by Coomassie brilliant blue staining and those containing $p21^{\text{mei3}}$ were pooled. A fraction of this material (300 μg) was applied to a 5.0-ml 15–35% glycerol gradient in 50 mM Tris pH 8.0, 50 mM NaCl, 3 mM MgCl_2 , 1 mM PMSF and centrifuged at 4 °C for 20 h at 49,000 r.p.m. in a SW50.1 rotor (Beckman). Molecular weight standards were sedimented in a parallel glycerol gradient. Twenty-one 0.2 ml fractions were collected from the bottom of the tube; 40 μl of each gradient fraction was applied to a 15% SDS-polyacrylamide gel and stained with Coomassie brilliant blue dye and 8 μl of each gradient fraction was added to a $p52^{\text{ran1}}$ kinase reaction (see Fig. 2).

ingly, the activity was also highly reduced in the $\text{ran1}^+ \text{O.P. mei3}^+ \text{O.P.}$ double-overexpressor, compared with the $\text{ran1}^+ \text{O.P.}$ strain (Fig. 3a), even though the same amount of $p52^{\text{ran1}}$ was immunoprecipitated from each cell lysate (Fig. 3b).

To test whether the low level of $p52^{\text{ran1}}$ autophosphorylation in the previous experiment is a direct or a secondary effect of mei3^+ expression, a strain of *E. coli* that expresses $p21^{\text{mei3}}$ at ~20% of total soluble cell protein⁹ was used. Addition of 1 μg of whole cell protein from this strain to $p52^{\text{ran1}}$ immunoprecipi-

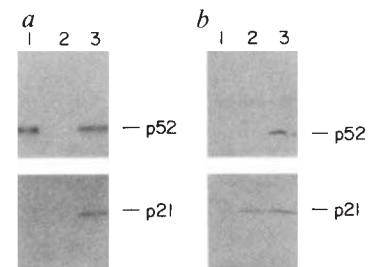


Fig. 5 Co-immunoprecipitation of $p52^{\text{ran1}}$ and $p21^{\text{mei3}}$. **a**, Lysates prepared from $\text{ran1}^+ \text{O.P. mei2}$ (SP800, lane 1), $\text{mei3}^+ \text{O.P. mei2}$ (SP799, lane 2), or $\text{ran1}^+ \text{O.P. mei3}^+ \text{O.P. mei2}$ (SP801, lane 3) strains were reacted with anti-p52 antibody. The immunoprecipitated material was subjected to SDS-PAGE and $p52^{\text{ran1}}$ (upper panel) and $p21^{\text{mei3}}$ (lower panel) were detected by probing immunoblots with R30 or N8 respectively. **b**, The same protocol was followed as in **a** except that polyclonal anti-p21 serum was used to form the initial immunocomplex.

tated from a $\text{ran1}^+ \text{O.P.}$ yeast strain (SP618), completely abolished protein kinase activity (Fig. 4a). The equivalent amount of protein from a control bacterium, that did not express $p21^{\text{mei3}}$, had no discernible effect (Fig. 4a). This suggests that $p21^{\text{mei3}}$ is a direct inhibitor of $p52^{\text{ran1}}$ kinase activity.

This observation was investigated further by purifying $p21^{\text{mei3}}$ from *E. coli* to virtual homogeneity by various chromatographic procedures (see Fig. 4 legend). The most highly purified fraction was subjected to velocity sedimentation through a 15–35% glycerol gradient and each fraction was assayed for inhibition of $p52^{\text{ran1}}$ kinase activity. The $p21^{\text{mei3}}$ sedimented at the expected velocity of a 21K monomer (Fig. 4b, upper panel) and the $p52^{\text{ran1}}$ inhibitory activity co-sedimented with $p21^{\text{mei3}}$ (Fig. 4b, lower panel). Thus, purified $p21^{\text{mei3}}$ directly inhibits $p52^{\text{ran1}}$. We estimated that ~200 ng of $p21^{\text{mei3}}$ were required to inhibit fully 100 ng $p52^{\text{ran1}}$ attached to protein-A sepharose beads.

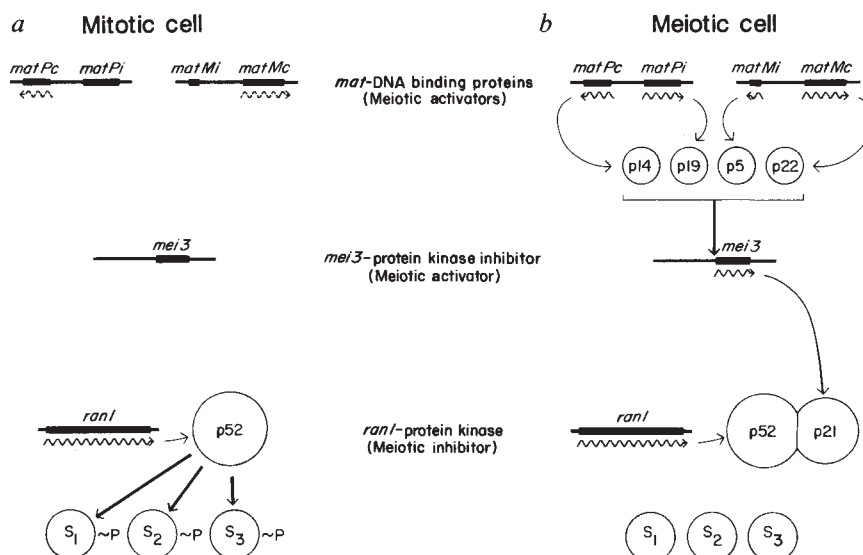
Protein complex

We considered several mechanisms by which $p21^{\text{mei3}}$ might inhibit $p52^{\text{ran1}}$ protein kinase. First, $p21^{\text{mei3}}$ might not strictly be a kinase inhibitor but might act instead as a phosphoprotein-phosphatase reducing the apparent incorporation of ^{32}P -ATP into $p52^{\text{ran1}}$ by dephosphorylating the protein. Alternatively, $p21^{\text{mei3}}$ might be a specific protease that degrades $p52^{\text{ran1}}$. These possibilities were largely excluded by the results of the following experiment: $p52^{\text{ran1}}$ that had been allowed to autophosphorylate *in vitro* in the presence of ^{32}P -ATP, was heat inactivated for 5 min at 65 °C, and then incubated with 1 μg purified $p21^{\text{mei3}}$ for 30 min. No effect of exposure of the ^{32}P - $p52^{\text{ran1}}$ to $p21^{\text{mei3}}$ was observed (data not shown).

Second, we investigated the possibility that $p21^{\text{mei3}}$ inhibits $p52^{\text{ran1}}$ by forming an inhibitory complex. If such a complex exists *in vivo* $p52^{\text{ran1}}$ and $p21^{\text{mei3}}$ might be co-immunoprecipitated, with either anti-p21 or anti-p52 sera, from a lysate prepared from a strain that expresses both proteins. Three strains were used, $\text{ran1}^+ \text{O.P. mei2}$ (SP800), $\text{mei3}^+ \text{O.P. mei2}$ (SP799) and $\text{ran1}^+ \text{O.P. mei3}^+ \text{O.P. mei2}$ (SP801). A whole cell lysate was prepared from cultures of each strain and was reacted with either anti-p21 or anti-p52 polyclonal sera. The immunocomplexes were precipitated with protein-A agarose, subjected to gel electrophoresis, followed by Western transfer and probed with both anti-p21 and anti-p52 monoclonal antibodies to reveal the abundance of each protein.

In a control experiment it was shown that $p21^{\text{mei3}}$ was not precipitated from the lysate of the $\text{mei3}^+ \text{O.P.}$ strain by anti-p52 serum, and nor was $p52^{\text{ran1}}$ precipitated from the $\text{ran1}^+ \text{O.P.}$ lysate by anti-p21 serum (Fig. 5a and b). This was expected as the two proteins show no obvious sequence homology^{9,11}.

Fig. 6 Model of meiotic control in *Schiz. pombe*. **a**, Mitotic cell. During vegetative growth of either haploid or diploid cells, the *mat-Pc* and *Mc* genes are weakly expressed but *Pi* and *Mi* are not expressed¹⁵. Consequently *mei3*⁺ is not expressed⁹. The *ran1*⁺ gene is expressed constitutively¹¹ and its 52K product, a protein kinase, is presumed to phosphorylate (heavy arrows) hypothetical substrates, S1, S2 and S3, which promote vegetative growth and inhibit meiosis. **b**, Meiotic cell. In an *h*⁺/*h*⁺ diploid, under nitrogen starvation conditions, each *mat* gene is transcriptionally induced¹⁵ and is presumed to be translated into proteins of the indicated molecular weights. Either directly or indirectly, the *mat* gene products induce transcription of the *mei3*⁺ gene⁹. The 21K *mei3*⁺ gene product binds to the p52^{ran1} protein kinase and inactivates it. This leads to dephosphorylation of S1, S2 and S3, and meiotic initiation.



However, each of the polyclonal sera individually immunoprecipitated both p21^{mei3} and p52^{ran1} from lysates prepared from the *ran1*⁺*mei3*⁺ double-overproducer strain (Fig. 5a, b). This result provides immunochemical evidence that p21^{mei3} and p52^{ran1} exist in a complex that survives exposure to mild detergent during the immunoprecipitation procedure, but not to the boiling SDS step that immediately precedes gel electrophoresis (see Fig. 5 legend). Thus, the association between p52^{ran1} and p21^{mei3} is stable but non-covalent. The structure of the p52^{ran1}/p21^{mei3} complex will be investigated in future studies.

Discussion

We obtained the following results. Meiosis is profoundly inhibited by overexpression of the *ran1*⁺ gene (Fig. 1, Table 1). This defect can be rescued and even overcompensated by simultaneous overexpression of *mei3*⁺, which induced haploid meiosis (Table 1). We showed that *ran1*⁺ encodes a protein kinase capable of autophosphorylation and residue Lys 47, the anticipated site of ATP binding, is essential for its enzymatic activity (Fig. 2). The p52^{ran1} immunoprecipitated from a *ran1*⁺*mei3*⁺ double-overproducer strain was complexed with p21^{mei3} and had strongly reduced autophosphorylation activity (Figs 3, 5). p21^{mei3} purified from *E. coli* is also inhibitory to p52^{ran1} activity *in vitro* (Fig. 4). These data suggest that inhibition of *ran1*⁺ activity is a necessary step in meiotic initiation, and demonstrate both genetically and biochemically that the *mei3*⁺ product, which is present only in meiotic cells⁹, is capable of inactivating *ran1*⁺.

Because of the low abundance of p52^{ran1} in wild-type cells we have used overexpressor haploid strains that mimic meiotic cells, rather than truly meiotic *h*⁺/*h*⁺ diploids. However, each of the above results is fully consistent with the following model for the control of entry into meiosis in *Schiz. pombe* (Fig. 6). Normally, meiosis occurs only in *h*⁺/*h*⁺ diploids under nutritional stress because these conditions allow expression of all four *mat* genes¹⁵. These genes are necessary for transcriptional induction of *mei3*⁺ (ref. 9). The *mei3*⁺ product forms a non-covalent complex with p52^{ran1} and inhibits its protein kinase activity (Figs 3–5). Inhibition of this kinase is sufficient to inhibit vegetative growth and to activate meiotic differentiation^{4–6}. It is assumed that *mei3* mutants are meiotically defective because p52^{ran1} cannot be inactivated in the absence of p21^{mei3}.

Because p52^{ran1} is a protein kinase this enzyme is expected to phosphorylate polypeptide substrates *in vivo* that promote vegetative growth and repress meiotic functions (Fig. 6). We have assayed only autophosphorylation activity and other physiological substrates have yet to be identified. However, no single

critical substrate of p52^{ran1} should necessarily be anticipated since this protein plays a very broad role in the vegetative cell cycle, cell growth, conjugation, meiosis and sporulation. For example, partial rather than full inactivation of the protein kinase in a *ran1ts* mutant does not provoke immediate meiosis and sporulation but instead derepresses sexual conjugation^{5,6}. Thus, *ran1ts* mutants cultured at a semipermissive temperature in rich medium arrest in the G₁ phase of the cell cycle and conjugate with cells of opposite mating-type⁶. This observation reflects the role played by *ran1*⁺ in the vegetative cell cycle and cannot be understood in terms of induction of meiotic or spore-specific functions. Furthermore, *mei3*⁺ cannot be involved in the regulation of p52^{ran1} kinase activity in haploid cells. We have previously suggested that an entirely separate pathway operates in vegetative cells that allows partial inactivation of *ran1*⁺ activity, and thus derepression of sexual conjugation, in response to nitrogen starvation⁶. Conjugation of course introduces both *mat-P* and *mat-M* into a single cell, allowing *mei3*⁺ to be expressed, inactivating p52^{ran1} and precipitating meiosis.

The mechanism of inhibition of p52^{ran1} by p21^{mei3} is of biochemical interest. The closest model may be provided by the so-called heat stable inhibitor of cyclic AMP-dependent protein kinase²⁶, although this protein shows no obvious sequence homology to p21^{mei3} (refs 9, 27). This 76 amino-acid residue polypeptide binds the free catalytic subunit of the protein kinase with very high affinity, in the presence of ATP, and competitively inhibits its enzymatic activity^{28,29}. The physiological role of the inhibitor is not fully clear at present.

The apparent mechanistic similarity of p21^{mei3} and the heat stable inhibitor of cAMP-dependent protein kinase may be more than simply formal. We have observed that gross overstimulation of cAMP-dependent protein kinase in fission yeast both inhibits normal meiosis and also by-passes the requirement for *ran1*⁺ for vegetative growth⁶. The role of cAMP-dependent protein kinase in *Schiz. pombe* is not well understood, but alterations in the activity of this enzyme may play a role in the transition from mitotic division to meiosis in *Sacc. cerevisiae*³⁰ and also in the amphibian *Xenopus*³¹. *ran1*⁺ protein kinase does not appear to be a cAMP-dependent protein kinase; except within the kinase domain the two proteins show no homology and *ran1*⁺ protein kinase is not sensitive to cAMP *in vitro* (ref. 11 and M.M., unpublished). However, a close interaction between the pathways regulated by these two protein kinases is anticipated.

Developmental biologists have long distinguished between the cryptic process of cell-type determination, which limits the range of possible developmental fates of a cell, and overt cellular differentiation, which often follows only after many further cell

divisions in response to an environmental stimulus. In the case of meiotic development in fission yeast these terms are equally applicable and can now be given some molecular basis. As in *Sacc. cerevisiae*, *Drosophila* and perhaps all eukaryotes, cell-type determination in *Schiz. pombe* is controlled by a class of master regulatory genes among which some, such as *mat-Pi* (ref. 15), contain a homoeobox sequence³². Cell differentiation, however, may be directly regulated by protein kinases. In fission yeast meiosis the missing link between these two levels of control is provided by the *mei3*⁺ gene⁹. It is transcriptionally regulated by *mat* and its product acts to inhibit the *ran1*⁺ protein kinase.

These results may open the way for an intriguing technical development. Given the conservation between yeast and verte-

brates of such fundamental housekeeping proteins as cAMP-dependent protein kinase^{33,34} and *cdc2*⁺ protein kinase^{35,36}, it is conceivable that homologues of *ran1*⁺ and *mei3*⁺ might exist in mammalian cells. *mei3*⁺ is a dominant meiotic activator, and could be used to induce meiotic haploidization in diploid tissue culture cells. If this result were achieved, formal genetic analysis might become a practical reality in such cells.

We thank Carmelita Bautista and Dianne Biederman for assistance with preparation of antibodies and Dave Greene and Jim Duffy for artwork. We also thank Rich Roberts and Bruce Stillman for critically reading the manuscript and Giulio Draetta for advice on biochemical techniques. Supported by NIH.

Received 9 December 1987; accepted 8 February 1988.

- Leupold, U. C. R. *Lab. Carlsberg Ser. Physiol.* **24**, 381-480 (1950).
- Beach, D. & Nurse, P. *Nature* **290**, 140-142 (1981).
- Bresch, C., Miller, G. & Egel, R. *Molec. gen. Genet.* **102**, 301-306 (1968).
- Iino, Y. & Yamamoto, M. *Molec. gen. Genet.* **198**, 416-421 (1985).
- Nurse, P. *Molec. gen. Genet.* **198**, 497-502 (1985).
- Beach, D., Rodgers, L. & Gould, J. *Curr. Genet.* **10**, 297-311 (1985).
- Beach, D., Nurse, P. & Egel, R. *Nature* **296**, 682-683 (1982).
- Beach, D. *Nature* **303**, 682-687 (1983).
- McLeod, M. & Beach, D. *EMBO J.* **6**, 729-736 (1987).
- Shimoda, C. & Uehira, M. *Molec. gen. Genet.* **201**, 353-356 (1985).
- McLeod, M. & Beach, D. *EMBO J.* **5**, 3645-3671 (1986).
- Egel, R. *Planta* **98**, 89-96 (1971).
- Egel, R. *Molec. gen. Genet.* **121**, 277-284 (1973).
- Egel, R. *Curr. Genet.* **8**, 199-203 (1984).
- Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. *EMBO J.* (in the press).
- Herskowitz, I. in *Specialized Cell Types; Their Use in Addressing Problems in Cell Biology* (ed. Hicks, J.) 625-656 (Liss, New York, 1986).
- Nasmyth, K. & Shore, D. *Science* **237**, 1162-1170 (1987).
- Egel, R. & Egel-Mitani, M. *Expl. Cell Res.* **88**, 127-134 (1974).
- Iino, Y. & Yamamoto, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2247-2451 (1985).
- Russell, P. *Nature* **301**, 167-169 (1983).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
- Zoller, M. & Smith, M. *Gene* **3**, 479-488 (1984).
- Weinmaster, G., Zoller, M. & Pawson, T. *EMBO J.* **5**, 1772-1779 (1985).
- Snyder, M., Bishop, J., McGrath, J. & Levinson, A. *Molec. cell. Biol.* **5**, 1772-1779 (1985).
- Lischwe, M. & Sung, M. *J. biol. Chem.* **152**, 4976-4980 (1977).
- Walsh, D. *et al. J. biol. Chem.* **246**, 1977-1985 (1971).
- Scott, J., Glaccum, M., Fischer, E. & Krebs, E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1613-1616 (1986).
- Whitehouse, S., Feramisco, J., Casnellie, J., Krebs, E. & Walsh, D. *J. biol. Chem.* **258**, 3693-3701 (1983).
- Whitehouse, S. & Walsh, D. *J. biol. Chem.* **258**, 3682-3692 (1983).
- Matsumoto, K., Uno, I. & Ishikawa, T. *Cell* **32**, 417-423 (1983).
- Maller, J. & Krebs, E. *J. biol. Chem.* **252**, 1712-1718 (1977).
- Gehring, W. *Science* **236**, 1245-1252 (1987).
- Shoji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 848-851 (1981).
- Toda, T., Cameron, S., Sass, P., Zoller, M. & Wigler, M., *Cell* **50**, 277-287 (1987).
- Draetta, G., Brizuela, L., Potashkin, J. & Beach, D. *Cell* **50**, 319-325 (1987).
- Lee, M. & Nurse, P. *Nature* **327**, 31-35 (1987).
- Beach, D., Piper, M. & Nurse, P. *Molec. gen. Genet.* **187**, 326-329 (1982).
- Spiczki, M. & Ferenczy, L. *Molec. gen. Genet.* **151**, 77-81 (1987).
- Gutz, H., Heslot, H., Leupold, U. & Loprieno, N. in *Handbook of Genetics* Vol. 1 (ed. King, R. C.) 395-446 (Plenum, New York, 1974).
- Losson, R. & Lacroute, F. *Cell* **32**, 371-377 (1983).
- Mitchison, M. *Meth. Cell Physiol.* **4**, 131-165 (1970).
- Laemmli, U. *Nature* **227**, 680-685 (1970).

LETTERS TO NATURE

The interaction of supernova 1987A with dense circumstellar gas

R. A. Chevalier

Department of Astronomy, University of Virginia, PO Box 3818, Charlottesville, Virginia 22903, USA

Although the progenitor of supernova 1987A apparently was a blue supergiant, there is theoretical and observational evidence for an earlier red supergiant phase. The fast wind from the blue star is expected to drive a shell into the slow, dense red supergiant wind. The physical properties of the shell depend on the wind characteristics and the temperature to which the gas drops behind the outer shock wave. Rayleigh-Taylor instabilities probably cause clumps to form in the shell. Observed narrow ultraviolet lines¹⁻⁵ are probably from shell gas that was photoionized by the initial radiation burst from the supernova. The light travel time effects on the evolution of the line fluxes, line profiles and spatial structure of the emission are described. Current high spectral resolution observations⁵ suggest a lower limit to the age of the shell of 5,000 yr and suggest a relatively low shell velocity. The narrow lines are predicted to shift to the red at a rate depending on the shell age. Observations of X rays below 10 keV by the Ginga satellite⁶ are difficult to interpret in terms of circumstellar interaction because of the time variability.

Chevalier and Fransson⁷ discussed the early observations of supernova 1987A related to circumstellar interaction. They showed that the early radio observations and lack of detectable X-ray emission were consistent with shock interaction with a

blue supergiant wind and predicted that radiative effects could include ultraviolet emission lines from gas ionized by the initial supernova burst and an infrared dust echo. Narrow ultraviolet lines have now been observed with the International Ultraviolet Explorer (IUE)¹⁻⁵. The likely site of the gas and dust emission is gas lost during an earlier red supergiant phase of the progenitor star, Sk - 69202. This was initially suggested on the basis of the Hertzsprung-Russell diagram for the Large Magellanic Cloud⁷, and subsequent stellar evolution calculations have shown that this is possible, although the time from the red supergiant phase to the explosion is not well determined. In a 20 $M_{\odot}$ model, Woosley⁸ finds a time of 2×10^4 yr and in a 15 $M_{\odot}$ model, 4×10^5 yr. These models have no mass loss. In a model with an initial mass of 20 $M_{\odot}$ and a final mass of 9.03 $M_{\odot}$ resulting from mass loss, Maeder⁹ finds a time of 10^5 yr. Of these models, the Woosley 20 $M_{\odot}$ model is especially promising as regards the progenitor properties and the light curve.

In its red supergiant phase, the Woosley 20 $M_{\odot}$ model has a radius of 4×10^{13} cm and a luminosity of $1 \times 10^5 L_{\odot}$. If the star were an OH/IR star, a mass-loss rate, $\dot{M}_r$, of $(3-4) \times 10^{-5} M_{\odot} \text{ yr}^{-1}$ would be appropriate¹⁰. If it were a normal red supergiant, the mass-loss rate could be almost a factor of 100 lower. The wind velocity, V_r , would be about 10 km s⁻¹, but could be in the range 5-20 km s⁻¹. The properties of the wind from the B3 star progenitor are a mass-loss rate, $\dot{M}_b$, of $\sim 5 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$ and a wind velocity V_b , of ~ 550 km s⁻¹ (refs 7, 11). The fast wind is expected to drive a shock front into the slow wind, sweeping up a shell of gas. In general, the shocked dense gas is radiative, whereas the shocked gas wind is non-radiative¹¹. Equations expressing conservation of mass and momentum in the dense shell and conservation of energy for the shocked fast wind lead to a cubic equation for the velocity,

V_s , of the shell

$$\dot{M}_r V_s \left(\frac{V_s^2}{V_r} - 2V_s + V_r \right) = \frac{1}{2+\alpha} \dot{M}_b V_b^2$$

where α is the fraction of the spherical volume of the shell occupied by the shocked fast wind and is likely to be close to 1. This equation, with $\alpha = 1$, is given by Kwok¹². The value of V_s derived in this way only applies asymptotically in time, that is after the fast wind has maintained its properties for roughly the doubling time of the shell. Because the stellar wind is likely to be less energetic during the evolution from the red to the blue phases, the actual shell velocity is expected to be between V_r and V_s . Table 1 shows some values of V_s for a blue star wind luminosity $L_b = 4.8 \times 10^{35}$ erg s⁻¹ and $\alpha = 1$. If L_b is reduced by a factor of 5, the values of V_s are reduced by 30–40%. During the transition between phases, shell acceleration is expected. A decline in the density of the red supergiant wind with radius also leads to shell acceleration. Acceleration of the dense shell by the hot, low density gas is Rayleigh–Taylor unstable and clumping within the shell is likely.

The basic properties of the shell are poorly constrained at present, so I use plausible values for the various parameters. The time scale $t_s = R_s/V_s$, where R_s is the shell radius, is about the age since the red supergiant phase. Some constraints of R_s come from the line intensities of the ultraviolet lines, although it is necessary to make certain assumptions about the initial ionizing burst from the supernova. Lundqvist and Fransson¹³ obtain $R_s \sim 2 \times 10^{18}$ cm. Then $V_s = 32(R_s/2 \times 10^{18} \text{ cm})(t_s/2 \times 10^4 \text{ yr})^{-1} \text{ km s}^{-1}$. The mass of the shell is

$$M_s = 1 \left(\frac{V_s}{V_r} - 1 \right) \left(\frac{\dot{M}_r}{5 \times 10^{-5} M_\odot \text{ yr}^{-1}} \right) \left(\frac{t_s}{2 \times 10^4 \text{ yr}} \right) M_\odot$$

and the hydrogen column density in the shell is $N_H = 2 \times 10^{19} (M_s/M_\odot)(R_s/2 \times 10^{18} \text{ cm})^{-2} \text{ cm}^{-2}$. The shell density depends on the temperature, T , that is attained in the post-shock layer during the presupernova evolution because the thermal pressure equals the shock ram pressure. Shocks that are fast enough to produce ionizing radiation have a temperature of about 5,000 K over the column density range 10^{18} – 10^{19} cm^{-2} ; the cooling column density to 100 K lies between 10^{19} and 10^{20} cm^{-2} for an ambient density less than 10^3 cm^{-3} (ref. 14). In the supernova 1987A shell, heating by the B3 star radiation field could be significant at the lower temperatures, or a magnetic field may limit the compression in the shell. If thermal pressure does dominate and the gas is neutral, the density in the shell is

$$n_H = 4.2 \times 10^3 \left(\frac{\dot{M}_r}{5 \times 10^{-5} M_\odot \text{ yr}^{-1}} \right) \left(\frac{V_r}{10 \text{ km s}^{-1}} \right)^{-1} \times \left(\frac{t_s}{2 \times 10^4 \text{ yr}} \right)^{-2} \left(\frac{T}{1,000 \text{ K}} \right)^{-1} \left(1 - \frac{V_r}{V_s} \right)^2 \text{ cm}^{-3}$$

Once the shell is photoionized by the radiation burst from the supernova, we have $n_e = 1.2 n_H$. The observed ratio of line intensities of the C III] doublet implies $n_e \approx 2 \times 10^4 \text{ cm}^{-3}$ (ref. 5), so that an initially cold post-shock layer may be required. The thickness of the dense shell is

$$\frac{\Delta R_s}{R_s} = 0.0014 \left(\frac{\dot{M}_r}{5 \times 10^{-5} M_\odot \text{ yr}^{-1}} \right) \left(\frac{V_r}{10 \text{ km s}^{-1}} \right)^{-1} \times \left(\frac{R_s}{2 \times 10^{18} \text{ cm}} \right)^{-2} \left(\frac{n_H}{2 \times 10^4 \text{ cm}^{-3}} \right)^{-1}$$

Although the high density implies a narrow shell, instabilities are likely to cause deviations from spherical symmetry.

The ultraviolet emission lines can yield a great deal of information on the circumstellar shell. They have already been used to demonstrate a large N overabundance in the gas³; here, we concentrate on the structure and motion of the circumstellar shell. Three simple cases are used to illustrate the light travel

Table 1 Shell velocity V_s (km s⁻¹)

V_r (km s ⁻¹)	$5 \times 10^{-6} \dot{M}_r$ ($M_\odot \text{ yr}^{-1}$)	10^{-4}
5	83	33
30	165	75

time effects on the flux evolution, velocity evolution, and spatial structure. Initially, the shell is assumed to be very thin, with radius R_s . The ionizing source is taken to be the initial burst of radiation at the time of shock breakout (at $t = 0$); the duration of this burst, Δt , is expected to be a few hours⁸. Case 1 assumes that the gas density is sufficiently high that the recombination time is less than Δt . The radiation burst then illuminates a thin ring on the shell from the point of view of a distant observer. If Θ is the angle between the line of sight to the supernova and a line between the supernova and the shell, then the position of the ring is given by $R_s(1 - \cos \Theta) = ct$, which follows from the additional light travel time. The radius of the ring as seen by the observer is $[ct(2R_s - ct)]^{1/2}$. The flux in a line is constant from $t = 0$ to $t = 2R_s/c$, when the flux vanishes. The observed line of sight velocity relative to the supernova is $V_1 = -V_s \cos \Theta = -V_s(1 - ct/R_s)$ for a narrow line with width $c\Delta t/t_s$, where $t_s = R_s/V_s$. Case 2 assumes that the density is sufficiently low that the recombination time is greater than $2R_s/c$. Now the observable surface of the shell increases with apparent radius $[ct(2R_s - ct)]^{1/2}$. The flux in a line increases linearly with time until $2R_s/c$, after which the flux remains constant until the recombination time. The observed velocity range is from $-V_s$ to $-V_s(1 - ct/R_s)$ until $2R_s/c$ and from $-V_s$ to $+V_s$ after that time. Thus the line centre has a velocity $V_1 = -V_s(1 - ct/2R_s)$ during the first phase. Case 3 is intermediate between cases 1 and 2. The gas recombines to the ion being observed at a time t_1 after being ionized, and recombines to a lower stage of ionization at time t_2 after ionization. In actuality the recombination is a gradual process, but we assume sharp transitions for purposes of illustration. Now, there is no observable flux until time t_1 when the surface of the shell first appears to be illuminated. The flux increases linearly as in case 2, until time t_2 . The observed velocity ranges from $-V_s$ to $-V_s[1 - c(t - t_1)/R_s]$. After t_2 , the flux remains constant, the velocity range is $-V_s[1 - c(t - t')/R_s]$, and the apparent radial range is $\{c(t - t')[2R_s - c(t - t')]\}^{1/2}$, where t' varies from t_1 to t_2 . After time $t_1 + 2R_s/c$, the flux declines linearly with time until $t_2 + 2R_s/c$ when the flux disappears. During this last phase, only a shrinking part of the receding shell is visible.

One general feature of the evolution is that the width of a line increases or decreases in relation to the flux of a line. Another is that the shell age can be expressed as

$$t_s = \frac{c}{\beta} \left(\frac{dV_1}{dt} \right)^{-1}$$

where $\beta = 1$ if the flux is constant, and $\beta = 2$ if there is a linear increase in flux. Although the above discussion has assumed a thin shell, this may break down, especially for case 1 where the emitting region is very narrow. The assumption breaks down for this case if $\Delta R/R_s > \Delta t/t_s$. If the line width is determined by the shell thickness, it is $ct\Delta R/(t_s R_s)$.

The ultraviolet lines from supernova 1987A were first detected in May 1987 and showed some flux increase through November 1987 (ref 1–5). It may not be possible to rule out a simple linear increase because $t = 0$, as in case 2, but it is probably more likely that the lines first brightened in May and later showed a more gradual increase, as expected in case 3 (ref. 13). High-resolution observations with the IUE on 25 November, 1987 set an upper limit to the width of the C III] and N III] lines of 30 km s^{-1} (full width at half maximum) (ref. 5). If the brightening from May

to November is attributed to an increasing fraction of the shell becoming visible because of light travel time effects, then

$$t_s = \frac{c(t - t_1)}{\Delta v} > \frac{c(0.5 \text{ yr})}{30 \text{ km s}^{-1}} = 5 \times 10^3 \text{ yr}$$

and $R_s > c(0.5 \text{ yr})/2 = 2.4 \times 10^{17} \text{ cm}$. In this model, the velocity of the emitting gas would extend to $-V_s$. But the average observed velocity of the lines is $+284 \pm 6 \text{ km s}^{-1}$ (ref. 5) and the highest velocity Ca II absorption component in the direction of supernova 1987A is $+295 \text{ km s}^{-1}$ (ref. 15). This suggests that the shell velocity is less than about $30\text{--}40 \text{ km s}^{-1}$, unless the supernova progenitor had an unusually high velocity relative to the surrounding gas.

Further high-resolution observations are warranted. The lines are predicted to shift to the red with time at a rate that is related to t_s . For $t_s = 2 \times 10^4 \text{ yr}$, dV_1/dt in the range $7.5\text{--}15 \text{ km s}^{-1} \text{ yr}^{-1}$ is expected. For a situation like case 3, the different ions have different time behaviours. As more highly ionized gas occurs closer to the ionizing flash, higher ionization stages appear relatively red-shifted. The N III and N V lines should be searched for this effect. Finally, it may be possible to spatially resolve the ultraviolet emission. For a distance of 50 kpc and $R_s = 2 \times 10^{18} \text{ cm}$, the angular diameter of the emission region grows to a maximum of 5.3 arc s. The combination of the evolution of line fluxes, line profiles, and spatial structure would give a detailed picture of the circumstellar gas. It is likely that the simple spherically symmetric model described here would break down and that asymmetric structure due to instabilities, non-spherical winds, or effects of nearby stars would become evident.

The circumstellar model described here is not compatible with the thermal X-ray model proposed¹⁶ to explain the low energy flux observed by the Ginga satellite⁶. In this model, $t_s = 10^2 (V_s/30 \text{ km s}^{-1})^{-1} \text{ yr}$ and the red supergiant wind extends in to a radius of $1.1 \times 10^{16} \text{ cm}$ to produce a rapid turn-on of thermal X-ray emission generated by shock interaction with the red supergiant wind. But the nature of the X-ray emission below 10 keV is still uncertain. Upper limits set by the TTM instrument on the Mir space station during August 1987¹⁷ may conflict with the Ginga results. Also, the observations by Ginga show strong variability on a time scale of weeks (Y. Tawara, personal communication). This is not naturally produced in the model of Masai *et al.*¹⁶. If a time of $2 \times 10^6 \text{ s}$ is identified with the isobaric cooling time of gas at $4.9 \times 10^7 \text{ K}$ (the temperature of a thermal fit to the Ginga observations: Y. Tawara, personal communication), a density of 10^9 cm^{-3} is required. A clump of this density would have to be far inside the supernova to be shocked to $4.9 \times 10^7 \text{ K}$, but it is unlikely that it would survive the effects of the outer supernova gas. It is more plausible to identify the time scale with adiabatic expansion and cooling of an inhomogeneity. The width of the shocked supernova gas, which is the expected source of X rays^{7,11}, divided by the sound speed of $4.9 \times 10^7 \text{ K}$ gas, can give the required time scale. But an inhomogeneity with an implausible degree of spherical symmetry is required. On a smaller scale, the interaction of a clump of supernova gas with a clump of circumstellar gas could probably model the emission, but is very *ad hoc*. Another possible problem is the lack of observable radio emission. We conclude that circumstellar interaction models for the X-ray emission are not promising.

Another observable manifestation of dense circumstellar matter is an infrared dust echo⁷. The expected luminosity is a strong function of R_s , however, and would be undetectable at present against the bright supernova light if $R_s = 2 \times 10^{18} \text{ cm}$. The signature of dust emission, either specific dust emission features or a clear rise to a blackbody peak at long wavelengths, has not yet been observed in spectra of supernova 1987A. Limits that can be set on R_s and predictions for infrared observations once the supernova fades are under study (R. T. Emmering and R.A.C., manuscript in preparation).

At the present time, ultraviolet line observations give the best

evidence for dense circumstellar gas and its properties. High spectral resolution observations should soon yield an estimate of t_s , the age of the dense shell, and should eventually yield the shell's radius and velocity. The existing information on the shell density and velocity already suggests a high rate of mass loss during the red supergiant phase. Light curve models for supernova 1987A⁸ indicate presupernova mass loss of $4\text{--}9 M_\odot$, and much of this mass may have been lost during the red supergiant phase.

I am grateful to C. Fransson and C. Sarazin for discussions. This research was supported in part by the NSF and NASA.

Received 31 December 1987; accepted 23 February 1988.

1. Wamsteker, W., Gilmozzi, R. & Cassatella, A. *IAU Circ. No.* 4410 (1987).
2. Kirshner, R. P. *et al. IAU Circ. No.* 4435 (1987).
3. Cassatella, A. in *Proc. ESO Workshop on SN1987A* (ed. Danziger, I. J.) 101–111 (ESO, Garching, 1987).
4. Kirshner, R. P. in *Proc. ESO Workshop on SN1987A* (ed. Danziger, I. J.) 121–131 (ESO, Garching, 1987).
5. Panagia, N. *et al. IAU Circ. No.* 4514 (1987).
6. Dotani, T. *et al. Nature* **330**, 230–232 (1987).
7. Chevalier, R. A. & Fransson, C. *Nature* **328**, 44–45 (1987).
8. Woosley, S. E. *Astrophys. J.* (in the press).
9. Maeder, A. in *Proc. ESO Workshop on SN1987A* (ed. Danziger, I. J.) 251–270 (ESO, Garching, 1987).
10. Herman, J. in *Mass Loss from Red Giants* (eds Morris, M. & Zuckerman, B.) 215–220 (Reidel, Dordrecht, 1985).
11. Chevalier, R. A. in *Proc. ESO Workshop on SN1987A* (ed. Danziger, I. J.) 481–494 (ESO, Garching, 1987).
12. Kwok, S. *Phys. Rep.* (in the press).
13. Lundqvist, P. & Fransson, C. in *Proc. ESO Workshop on SN1987A* (ed. Danziger, I. J.) 495–502 (ESO, Garching, 1987).
14. McKee, C. F. & Hollenbach, D. J. *A. Rev. Astr. Astrophys.* **18**, 219–262 (1980).
15. Vidal-Madjar, A. *et al. Astr. Astrophys.* **177**, L17 (1987).
16. Masai, K., Hayakawa, S., Itoh, H. & Nomoto, K. *Nature* **330**, 235–236 (1987).
17. Sunyaev, R. *et al. Nature* **330**, 227–230 (1987).

Gamma-rays and X-rays from SN1987A

Mark D. Leising

E. O. Hulburt Center for Space Research,
Naval Research Laboratory, Washington, DC 20375, USA

Detections of γ -rays¹ and X-rays^{2,3} from Supernova 1987A, together with its exponential light curve⁴, have dramatically confirmed theoretical predictions made two decades ago^{5–7}. We discuss what these observations, in August and September 1987, can already tell us about the supernova. Assuming that all of the high-energy photons originate from ^{56}Co decay, the escaping photon intensities at various energies are related by the characteristics of the scattering medium and the physics of the scattering processes. Considering only these factors, we conclude that the total mass of ^{56}Co implied by the light curve is not uniformly shielded by a thick envelope. Extensive 'mixing' of ^{56}Co (initially ^{56}Ni) and other heavy elements through the supernova envelope might explain the observations. It is likely that a small fraction of the ^{56}Co mass implied by the light curve has been removed by some mechanism to nearly thin regions.

Many authors have discussed the importance of, and prospects for, observations of γ -ray lines and the associated hard continuum from supernovae in general^{6,8–10}, and SN1987A in particular^{11–17}. Such observations not only verify theories of supernova nucleosynthesis but can also provide important clues to the structure and content of the ejected envelope. The detections of γ -ray lines and X-ray continua, although earlier than expected, might begin to yield these clues or might portend unforeseen complexities. The studies of SN1987A cited above did not predict the reported detections, but variations in the supernova envelope models that serve as input to the photon transfer calculations might accommodate the observations. Here we offer a simple attempt to understand the observations without reference to any particular models of supernova atmospheres.

We take the reported γ -ray and X-ray observations at face value, although the possibility of undiscovered systematic errors remains.

The ideas are straightforward. The Compton scattering cross-section decreases with energy, so the free escape of higher-energy lines from a scattering medium is enhanced compared to lower-energy lines. The continuum is produced by multiple scatterings of line photons at various angles. How far down in energy the photons are scattered before escape depends on the number of scatterings, and therefore on the depth of the scattering medium. The lower-energy continuum is suppressed by photoelectric absorption on heavy elements, so the composition of the scattering material will also affect the intensity of the lower-energy continuum relative to the lines.

We consider the transfer of the photons in two concentric, homogeneous, spherical shells defined by their radial Thomson scattering thicknesses (τ_T). The inner shell has ^{56}Co and heavy elements throughout and the outer shell consists of material with solar abundances. This is a reasonable, although greatly simplified, description of a supernova envelope for this problem. We perform a Monte Carlo calculation, treating only Compton scattering (employing the Klein-Nishina differential cross-section) and photoelectric absorption, to monitor the propagation and emergence of the photons. The radii of the two shells are given by a reasonable estimate of the velocities in the supernova envelope, but the dependent geometric effects are small. We first examine the effect of various thicknesses of the overlying envelope which has no line emissivity and low bound-free opacity.

Our calculated emerging spectra are consistent with those for envelopes of (we estimate) similar thicknesses in published models (refs 15-17). Figure 1 shows the ratios of the intensities of the lines from ^{56}Co decay at 1,238 keV and 2,599 keV to that at 847 keV, as a function of the Thomson optical depth of the outer shell, for a thin ($\tau_T = 1$) inner shell. Also shown is the fraction of all emitted 847-keV photons which escape unscattered. Included for comparison are the measured 1,238 keV to 847 keV line ratio and inferred 847-keV escape fraction¹. Figure 2 shows the integrated intensities of two continuum bands (25-50 keV and 100-200 keV) relative to the 847-keV line intensity for the same model. These bands, arbitrarily chosen for ease of illustration, are within the measured continuum range. The 25-50 keV emission is very sensitive to elemental abundances through photoelectric absorption. The higher-energy band is more directly related to the lines by the depth of the scattering medium.

If $0.07 M_\odot$ of ^{56}Ni was produced initially⁴, the 847-keV line flux detected by the Solar Maximum Mission (SMM) spectrometer¹ could result from the escape of only 1.3% of the total 847-keV luminosity of the ^{56}Co which remains 200 days after the explosion. From Fig. 1 we see that when the escape fraction of the 847-keV photons is this small, the ratios of the 1,238-keV and 2,599-keV lines to the 847-keV line are 1.2 and 0.8, respectively. The reported¹ ratio of 1,238-keV flux to that at 847 keV is 0.6 ± 0.25 . In addition, no line is apparent in the SMM subtracted spectrum at 2,599 keV at an intensity approaching 0.8 of the measured 847 keV intensity (S. M. Matz, personal communication). Thus the γ -ray line ratios alone suggest that it is unlikely that SMM is detecting the total mass of ^{56}Co through a thick spherical envelope.

Now consider the X-ray observations in relation to the γ -ray line intensities, taking the measurements as concurrent. Instruments on the Mir spacecraft detected a continuum from 20 to 300 keV with a power-law index of ~ -1.4 (ref. 2). Integrating that power law yields integral intensities of $\sim 1.7 \times 10^{-3} \text{ cm}^{-2} \text{ s}^{-1}$ in the 25-50 keV range and $\sim 10^{-3} \text{ cm}^{-2} \text{ s}^{-1}$ from 100 to 200 keV. The ratios of the intensities in these bands to that of the 847-keV line detected by SMM are 1.7 and 1.0 respectively. From Fig. 2 we see that when the 847-keV intensity is attenuated to 1.3%, the continuum bands are 12 and 7 times as intense as that line.

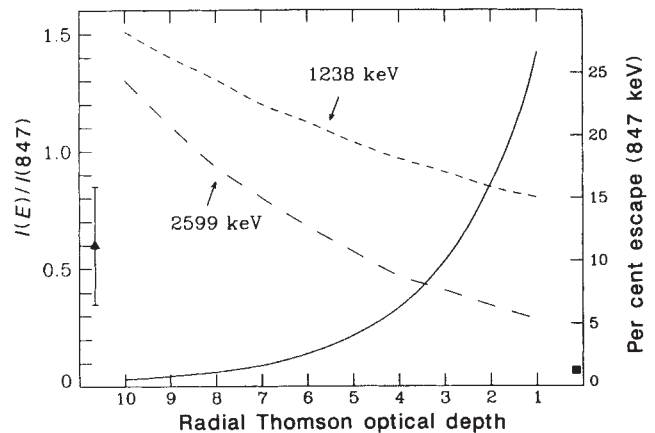


Fig. 1 Monte Carlo calculation of emerging γ -ray line intensities for strongest ^{56}Co lines. Solid line, the per cent escape of 847-keV photons; small dashes, the ratio of emerging 1,238 keV to 847 keV fluxes; long dashes, the ratio of emerging 2,599 keV to 847 keV fluxes; all plotted versus radial Thomson optical depth of overlying hydrogen-rich shell (for $\tau_T = 1$ inner shell). Triangle near left axis shows ratio of 1,238 keV line to 847 keV line measured by SMM¹. Square near right axis shows measured 847 keV escape fraction¹, assuming production of $0.07 M_\odot$ of ^{56}Ni 200 days earlier.

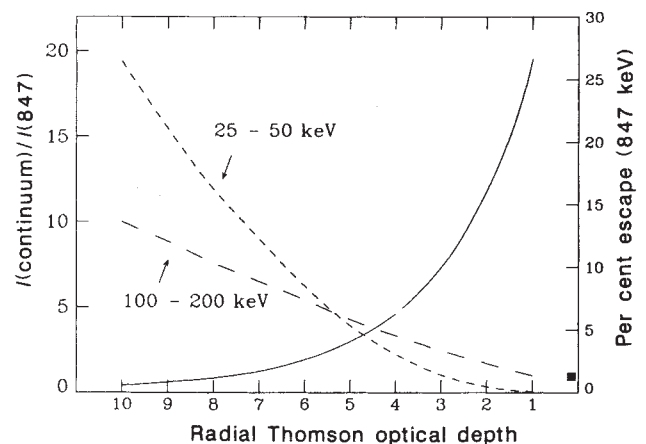


Fig. 2 Similar to Fig. 1, but ratios of integral intensities in two continuum bands to 847 keV line are shown. Small dashes, 25-50 keV continuum; long dashes, 100-200 keV continuum.

That is, if there were enough overlying envelope material to scatter away such a large fraction of the 847-keV line, the continuum would be significantly more intense than observed. This conclusion can also be inferred from the results of the more detailed calculations of the X-ray emission from specific supernova envelope models referenced above. Note that the observed X-ray spectrum² is roughly compatible with an envelope this thick, but the continuum intensity is too low relative to the lines.

The 25-50 keV continuum intensity relative to the 847-keV line intensity suggests an overlying envelope of thickness $\tau_T = 2-3$, depending somewhat on the inner shell thickness and elemental abundances. For a thinner envelope, not enough scatterings occur in the low-opacity material to produce photons at these energies, whereas a thicker envelope scatters many more photons into this band than are in the line. The small ratio of the >100 keV continuum intensity to that of the 847 keV line suggests that the thickness of any envelope above the ^{56}Co is even less, with $\tau_T \leq 1$. At 200 days after the outburst, when the supernova ejecta should be very thick, such a thin overlying

envelope might result from the mixing of the inner, high-metallicity material outward. The redistribution of the initially stratified supernova ejecta has been invoked to explain the supernova light curve and the earlier than predicted X-ray detection¹⁵⁻¹⁷.

For models with an outer shell of thickness $\tau_T = 2$, we find that an inner shell (a self-absorbing source region) of thickness $\tau_T \approx 25$ is required to reduce the escaping intensity at 847 keV to the observed value. Although this is a reasonable total thickness for several $M_\odot$ of material moving with velocities of a few thousand km s⁻¹, homogeneous mixing of the entire inner envelope so near the surface seems rather extreme, and whether the energy contained in the nickel and cobalt is sufficient to drive such extensive redistribution is questionable. Such a model is in fair agreement with the observations, but the calculated continuum is slightly harder (roughly an $E^{-1.1}$ power-law), and more intense relative to the 847-keV line, than observed², and photoelectric absorption essentially eliminates photons of energy <35 keV. This last point would seem to be incompatible with the Ginga observations³, but the results of the calculation at the lower energies are quite sensitive to the treatment of metallicity. If other heavy elements are not mixed quite as extensively as the ⁵⁶Co, the photoelectric cutoff moves to lower energies and that discrepancy with the observations is avoided.

Some such nonhomogeneous mixing can probably explain all the observations, and not require more energy than is contained in the radioactive material. If a small fraction of the ⁵⁶Co was carried to very thin regions, the observed line ratios would obtain essentially their laboratory values, whereas most or all of the continuum emission would originate under more envelope. The absence of a significant increase in the line fluxes with time also suggests that the ⁵⁶Co is not uniformly distributed (in mass coordinate) in the outer envelope.

In an extreme version of this model the lines might be emitted by material ejected completely out of the envelope, as in a jet. Ejection of core material has been proposed^{18,19} as the cause of the companion object²⁰ with radioactivity as a possible energy source. The implied initial ⁵⁶Co masses in that object (from its luminosity) and in the source of the 847-keV line (if it were completely thin at 200 days) are similar. But the companion is probably not the source of the γ -rays, because its relativistic velocity, if it was ejected by the supernova, would shift the line energies.

The subsequent evolution of the high-energy fluxes could reveal the degree of inhomogeneity of the mixing. If real, the possible decline¹ of the 847-keV line intensity would argue for a small, nearly thin mass of ⁵⁶Co as its source. If it is so, the γ -rays might return at detectable levels as the envelope thins, and all the high-energy emission might be compatible with published models with moderate degree of envelope mixing. Future line detections of sufficient significance might show relatively short timescale variations of line fluxes, energies and widths, indications of very nonhomogeneous mixing. If almost the entire envelope is uniformly mixed, the line fluxes should be observed to increase slowly into early 1988, the <50 keV X-ray flux should decline as the small overlying envelope thins, and the >100 keV X-ray continuum should persist for quite some time. Such a model would also predict that the lines would be blue-shifted and broadened, both by $\sim 1\%$. Balloon observations with a high-resolution spectrometer of a 3σ excess at 847 keV (W. G. Sandie *et al.*, preprint), consistent with the SMM observations, indicate that that line is at its rest energy, although the line energy is not well determined. This is compatible with the emitting material moving roughly perpendicular to the line of sight, a situation most easily explained by a small mass of ⁵⁶Co emerging from the massive envelope.

This work was performed while I held a National Research Council Research Associateship at the Naval Research Laboratory. I thank Gerry Share and Steve Matz for useful discussions and Don Clayton for his encouragement of this work.

Received 28 January; accepted 23 February 1988.

1. Matz, S. M. *et al. Nature* **331**, 416-418 (1988).
2. Sunyaev, R. *et al. Nature* **330**, 227-229 (1987).
3. Dotani, T. *Nature* **330**, 230-231 (1987).
4. Catchpole, R. M. *et al. Mon. Not. R. astr. Soc.* (in the press).
5. Bodansky, D., Clayton, D. D. & Fowler, W. A. *Astrophys. J. Suppl. Ser.* **16**, 299-371 (1968).
6. Clayton, D. D., Colgate, S. A. & Fishman, G. J. *Astrophys. J.* **155**, 75-82 (1969).
7. Colgate, S. A. & McKee, C. *Astrophys. J.* **157**, 623-643 (1969).
8. Clayton, D. D. *Astrophys. J.* **188**, 155-157 (1974).
9. Woosley, S. E., Axelrod, T. S. & Weaver, T. A. *Comments Nucl. Part Phys.* **9**, 185-197 (1981).
10. Gehrels, N., Leventhal, M. & MacCallum, C. J. *Astrophys. J.* **322**, 215-233 (1987).
11. McCray, R., Shull, J. M. & Sutherland, P. *Astrophys. J.* **317**, L73-L77 (1987).
12. Chan, K. W. & Lingenfelter, R. E. *Astrophys. J.* **318**, L51-L55 (1987).
13. Gehrels, N., MacCallum, C. J. & Leventhal, M. *Astrophys. J.* **320**, L19-L22 (1987).
14. Ebisuzaki, T. & Shibazaki, N. *Astrophys. J.* (in the press).
15. Itoh, M. *et al. Nature* **330**, 233-235 (1987).
16. Pinto, P. & Woosley, S. E. *Astrophys. J.* (in the press).
17. Shull, J. M. & Xu, Y. in *Proc. George Mason Workshop on SN1987A* (in the press).
18. Piran, T. *Nature* **330**, 28 (1987).
19. Goldman, I. *Astr. Astrophys.* **186**, L3-L4 (1987).
20. Nisenson, P. *et al. Astrophys. J.* **320**, L15-L18 (1987).

X-ray emission from the ring nebula NGC6888

N. G. Bochkarev

Sternberg State Astronomical Institute, Universitetskij prosp. 13, Moscow V-234, USSR

Soft X-ray emission from the ring nebula NGC6888, formed by the stellar wind of the central Wolf-Rayet star HD192163, has been found after thorough processing of Einstein observation data. A new class of X-ray sources, predicted some 20 years ago, has thus been discovered. The spectral shape agrees completely with theoretical predictions, but the emission flux is more than an order of magnitude less than the predictions.

The nebula NGC6888 around the WN6 star HD192163 is a prototype of a broad class of ring nebulae around stars with strong stellar winds^{1,2}. They play an important role in the formation of the structure of the interstellar medium, but their origin has not yet been completely resolved. Three possibilities are discussed: (1) ambient interstellar gas is swept up by strong stellar wind; (2) a shell of material is expelled during the evolution of the central star; and (3) a combination of processes (1) and (2)³.

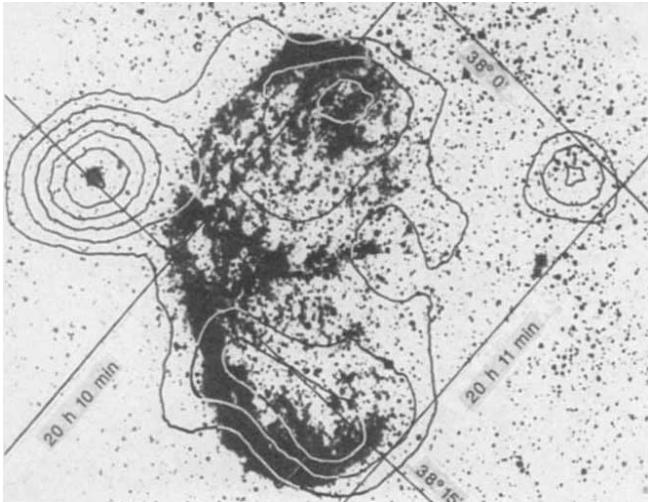
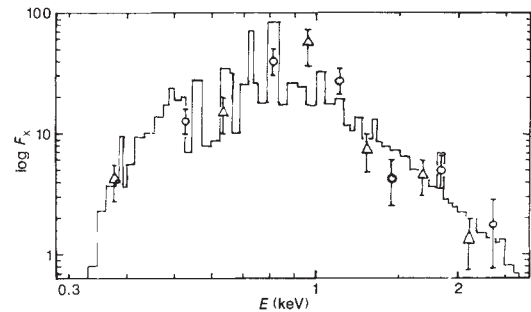
A two-layered shock wave is formed in the case of strong stellar wind⁴. Each layer evolves through energy-conserving and momentum-conserving phases, similarly to supernova remnants⁵. Like the majority of nebulae blown by stellar winds, the decelerated layer of the stellar wind of NGC6888 is observed in the adiabatic phase and the layer of compressed interstellar matter is observed in the radiative phase⁶. The decelerated wind layer must be a source of soft X-rays⁴. The structure of such a layer has been calculated in the case of a nebula expanding into a homogeneous interstellar medium^{7,8}. At the same time, statistics on such nebulae have been used⁹ to suggest that the stellar wind should rather effectively leak through the gaps of the regular shell structure. In such a case the X-ray flux may be substantially weakened.

The results of the first attempts to detect X-rays from ring nebulae of this type were negative. The Einstein data did not indicate any X-ray flux from Sh308 and RCW58 (ref. 10). Recent Exosat results on NGC6888 have also proved to be negative¹¹.

Calculating X-ray fluxes from ring nebulae using the stellar wind-interstellar medium interaction model¹² has shown that for two nebulae (NGC6888 and S119) the flux is at the IPC (imaging proportional counter) detectability level of the Einstein observatory. The NGC6888 region was observed with the IPC on 29 April 1979. The effective observation time was 10,756.8 s.

Table 1 X-ray sources detected by IPC centred on the NGC6888 nebula

Coordinates of centre (1950.0)			Radius, arc min	Signal		kT (keV)	Spectral parameters		Source
Source	RA	Dec		count	confidence		N_{H} (cm ⁻²)	$F_{\text{x}}(0.15\text{--}4\text{ keV})$ (erg cm ⁻² s ⁻¹)	
1	20 h 09 m 36 s	38° 14.9′	3.00	152	7.9σ	0.17–0.35	≤10 ²⁰	(3–4) × 10 ⁻¹³	HD192020
2	20 h 10 m 40 s	38° 15.0′	5.33	320	9.1σ	0.23–0.75	3 × 10 ²¹	(5.5–6.5) × 10 ⁻¹³	NGC6888 + HD192182
3	20 h 10 m 08 s	38° 7.5′	4.17	208	8.2σ	0.28–0.80	3 × 10 ²¹	(4–5) × 10 ⁻¹³	NGC6888
4	20 h 10 m 52 s	38° 1.0′	3.00	77	4.7σ	≥1	≤3 × 10 ²²	(1.5–3) × 10 ⁻¹³	point source

**Fig. 1** Isophotes of soft X-rays emission from NGC6888, superimposed onto the optical image of the ring nebula. The contours correspond to 1, 2, 3, 4, 5 and 6 counts arc min^{-2} (10^4 s^{-1}) of the Einstein IPC. Coordinates are for the epoch 1950.0.**Fig. 2** Comparison (in relative units) of the observed NGC6888 X-ray spectrum (Δ , $\circ$) with the calculated spectrum (histogram) of a nebula blown up by a stellar wind with the central zone temperature of the decelerated stellar wind $T = 10^7 \text{ K}$ and with the column density of absorbing interstellar medium $N_H = 10^{21.5} \text{ cm}^{-2}$. In the northern part of the nebula X-ray image (source 2 in Table 1) the soft tail of the spectrum is not shown because it is determined by the star HD192182, which is behind the sources; the remaining spectrum is indicated with circles. In the southern part (source 3) the observed spectrum is indicated with triangles. 1σ confidence intervals are shown.

The IPC was centred on HD192163 in the nebula, and the field was $\sim 1^\circ \times 1^\circ$. The optical image of the nebula was $12' \times 17'$. The signal was ~ 970 counts (13 standard deviations) in excess of the background. The image contains four components, two point-like components (1 and 4 in Table 1) and two extended components seen in the optical image of the nebula near the ends of its major axis (Fig. 1).

The data were processed by the standard method¹³ for each of the four sources separately. Two procedures for background subtraction were checked. First, source-free areas of the image of the same field were taken as background, and second, for comparison the data from a nearby field, lacking any marked sources and accumulated for 33,000 s, were processed as background. The results proved to be similar.

The IPC observations yield information about source spectra. The processing was done using the data on 8-9 spectral intervals in the $0.1 < E < 4 \text{ keV}$ energy range. The spectra were approximated either by power laws or by the emission spectrum of an optically thin plasma with 'normal' cosmic elemental composition, according to Raymond's model¹⁴.

The results are summarized in Table 1, which presents the coordinates of the centres of the circles over which the signal (radiation excess over background) was integrated, the radii of the circles, the value of the signal from the object (in counting rates of the detector and in units of dispersion of counts from background) the source temperature interval (kT) at a 90% significance level calculated according to ref. 15, the adopted value of column density of absorbing interstellar medium N_H , and the observed flux $F_x(0.15-4 \text{ keV})$.

All four sources were detected at a high significance level. Their spectra can be approximated by both power law and the spectrum of optically thin plasma, with approximately equal values of squared-deviation sums. Except for source 4, however, the power-law approximation gives very steep spectra (the slope in photon number is $\alpha > 3.5$). In this case a thermal spectrum is more plausible. Source 4 alone has a hard spectrum $kT > 1 \text{ keV}$, or $0.5 < \alpha < 2$. From the observations it follows that $N_H < 3 \times 10^{22} \text{ cm}^{-2}$ for source 4. The source could be identified with neither bright stars, nor known quasars, nor known active galactic nuclei. Source 1 found earlier on Exosat has been identified with the nearby (25-40 pc) main-sequence star G5-G8 HD192020. A value $N_H < 10^{19}-10^{20} \text{ cm}^{-2}$ should be expected for this star^{16,17}.

According to ref. 1, the interstellar optical absorption throughout NGC6888 is uniform, and approaches $A_v = 1.6 \text{ mag}$. The column density N_H for the nebula (sources 2 and 3 in Table 1) has been inferred from the correlation between colour excess and N_H (see ref. 18 for example) for interstellar medium.

Nebulae formed by stellar winds are not isothermal^{7,8}. The temperature drop and rise in density from the centre to the boundary of decelerated stellar wind result in an enhancement of the low-energy tail of the spectrum, distorting its form substantially¹⁹. The kT data obtained by standard processing of spectra therefore yield only a certain weighted-mean temperature. Comparison of the observed X-ray spectral distribution with that predicted^{12,19} for the given type of nebulae shows excellent agreement (Fig. 2), whence it follows that the temperature of the central part of the decelerated stellar wind is

$(8-10) \times 10^6$ K, similar to its predicted value¹².

The X-ray emission from the northern part of NGC6888 (source 2) includes an admixture of a soft component which seems to be due to the weak background source (the G5 star HD192182) whose X-ray spectrum resembles the spectrum of source 1. The flux $F_x \approx 10^{-13}$ erg cm⁻² s⁻¹ from this star was estimated from the soft ($E < 0.3$ keV) X-ray excess in source 2 compared with source 3, and from the amount of energy incident on the IPC beam centred on HD192182 (RA = 20 h 10 min 24 s, dec. = 38°17.5'). No emission from the nebula central star HD192163 was noted. The flux from the star is $F_x < 5 \times 10^{-14}$ erg cm⁻² s⁻¹.

After subtracting the HD192182 radiation, the measured X-ray flux from NGC6888 appeared to be F_x (0.2–3.0 keV) = 10^{-12} erg cm⁻² s⁻¹ with a 12–15% uncertainty (at 1 σ level). This value agrees with the upper limit inferred from Exosat observations¹¹ $F_x(0.05-2.0 \text{ keV}) < 10^{-12}$ erg cm⁻² s⁻¹ and 10–20 times as low as the predicted value^{11,12}.

The complete agreement with the calculated spectral form seems to indicate that the interaction of the stellar wind of the central star with the interstellar medium is the dominant source of soft X-rays from NGC6888. The low flux level F_x may arise from the leakage of a significant fraction of stellar wind between interstellar medium inhomogeneities⁹, although it may also be significant that the object is young (25,000 years^{6,12}), so ionization equilibrium and hence equilibrium radiation have not yet been established¹⁹. The ionization is higher than the equilibrium value, so a deficit of radiation relative to the equilibrium spectrum is suggested.

At present the mass blown by the stellar wind, $\sim 0.5 M_\odot$, is ~ 10 times less than the mass of compressed interstellar gas⁶. This difference can explain the enhanced nitrogen and helium

abundances^{2,20} and shows that the observed velocities are not the original ejection velocities but associated with acceleration of the interstellar gas by the stellar wind. It means that for homogeneous interstellar matter, the X-ray flux would be close to that predicted by Bochkarev and Lozinskaya¹² and Bochkarev¹⁹. The interstellar gas or stellar wind is therefore expected to be inhomogeneous, in accord with hydrodynamical calculations²¹.

I thank Dr D. E. Harris and C. Stern for their assistance in processing data, and M. Toropova for help in preparing the figures.

Received 26 October 1987; accepted 5 January 1988.

1. Wendker, H. J., Smith, L. F., Israel, F. P., Habing, H. J. & Dickel, H. R. *Astr. Astrophys.* **42**, 173–185 (1975).
2. Kwitter, K. B. *Astrophys. J.* **245**, 154–162 (1981).
3. Lozinskaya, T. A. *Astrophys. Space Sci.* **87**, 313–331 (1982).
4. Pikel'ner, S. B. & Shcheglov, P. V. *Soviet Astr.* **12**, 757–764 (1968).
5. Shklovsky, I. S. *Supernovae* (Nauka, Moscow, 1976).
6. Lozinskaya, T. A. *Supernovae and Stellar Winds: Interaction with the Interstellar Medium* (Nauka, Moscow, 1986).
7. Castor, J., McCray, R. & Weaver, R. *Astrophys. J.* **200**, L107–L110 (1975).
8. Weaver, R., McCray, R., Castor, J., Shapiro, P. & Moore, R. *Astrophys. J.* **218**, 377–395 (1977).
9. Lozinskaya, T. A. *Sov. Astr. Lett.* **9**, 247–249 (1983).
10. Moffat, A. F. J., Firmani, C., McLean, I. S. & Seggewiss, W. in *IAU Symp. 99: WR Stars: Observations, Physics, Evolution* (eds DeLoore, C. W. H. & Willis, A. J.) 577–581 (Reidel, Dordrecht, 1982).
11. Kähler, H., Ule, T. & Wendker, H. J. *Astrophys. Space Sci.* **135**, 105–109 (1987).
12. Bochkarev, N. G. & Lozinskaya, T. A. *Soviet Astr.* **29**, 60–65 (1985).
13. Einstein Observatory (HEAO-B): User's Manual. CFA/HEA 80–320 (1980).
14. Raymond, J. C. & Smith, B. W. *Astrophys. J. Suppl. Ser.* **35**, 419–439 (1977).
15. Avni, Y. *Astrophys. J.* **210**, 642–646 (1976).
16. Paresce, F. *Astr. J.* **89**, 1022–1037 (1984).
17. Bochkarev, N. G. *Astrophys. Space Sci.* **138**, 229–302 (1987).
18. Spitzer, L. Jr. *Physical Processes in the Interstellar Medium* (Wiley, New York, 1978).
19. Bochkarev, N. G. *Soviet Astr.* **29**, 509–515 (1985).
20. Parker, R. A. R. *Astrophys. J.* **224**, 873–884 (1978).
21. Rózycka, M. & Tenorio-Tangle, G. *Astr. Astrophys.* **147**, 202–226 (1985).

High-temperature shock formation of N₂ and organics on primordial Titan

Christopher P. McKay*, Thomas W. Scattergood†, James B. Pollack‡, William J. Borucki‡ & Hilde T. Van Ghysegem‡

* Solar System Exploration Branch, NASA Ames Research Center, Moffett Field, California 94035, USA

† State University of New York, Stony Brook, New York 11794, USA

‡ NASA Ames Research Center, Moffett Field, California 94035, USA

Titan's atmosphere is composed primarily of N₂ with a little methane and other organic molecules. But theoretical models suggest that the initial form of nitrogen in Titan's atmosphere may have been NH₃. We have investigated the possible importance of strong shocks produced during high-velocity impacts accompanying the late stages of accretion as a method for converting NH₃ to N₂. To simulate the effects of an impact in Titan's atmosphere we have used the focused beam of a high-power laser, a method that has been shown to simulate shock phenomena. For mixtures of 10%, 50% and 90% NH₃ (balance CH₄) we obtained yields of 0.25, 1, and 6×10^{17} molecules of N₂ per joule, respectively. We also find that the yield of HCN is comparable to that for N₂. In addition, several other hydrocarbons are produced, many with yields in excess of theoretical high-temperature-equilibrium models. The above yields, when combined with models of the satellite's accretion, result in a total N₂ production comparable to that present in Titan's atmosphere and putative ocean.

Studies of the formation of the giant planets indicate that Titan may have formed in a nebula disk spun out of the outer envelope of Saturn during its early contraction phase^{1,2}. In this case, a given parcel of gas that contributed to the final satellite

would have traversed regions of sufficiently high pressure and temperature to insure that almost all of the C was in the form of CH₄ and that almost all of the N was in the form of NH₃³. In addition, if N₂ was the dominant form of nitrogen in the Saturn nebular disk, much more CO than CH₄ would have been incorporated into Titan⁴, contrary to the observed composition of its atmosphere. As a result, it is possible that Titan accreted most of its nitrogen in the form of NH₃.

Of interest therefore is the mechanism whereby an initial NH₃–CH₄ atmosphere could have been altered to produce the present N₂-rich mixture. Atreya *et al.*⁴ suggested that the conversion of NH₃ to N₂ can be achieved by solar ultraviolet light; but it may have been difficult to photodissociate NH₃ on Titan if the tropopause cold trap greatly reduced the ammonia abundance in the stratosphere, as it does today, or if the solar nebula, Saturn's nebula, or Titan's upper atmospheric haze were present and had significant ultraviolet opacity during the satellite's accretion, when high surface temperatures may have occurred.

We have investigated an alternative method for converting NH₃ to N₂: shocks produced during high-velocity meteor impacts. Shock production of chemical species has been considered theoretically for NO on the present Earth^{5–8}, for the early Earth^{9,10} and for Titan-like gas mixtures^{11,12}. We are considering those impacts that heat the entrained air to high pressures and temperatures, of the order of several thousand degrees. In the high-temperature region, chemical reactions proceed swiftly and the various chemical species come rapidly to their thermodynamic equilibrium concentrations. But as the shock cools below $\sim 2 \times 10^3$ K, by expansion and radiative heat losses, the molecules formed do not stay in equilibrium with the cooling gases, because of the increasingly slow reaction rates, so they are left in concentrations characteristic of a high-temperature equilibrium⁶.

Considerable shock energy would have been available on Titan during the later stages of accretion. To produce a high

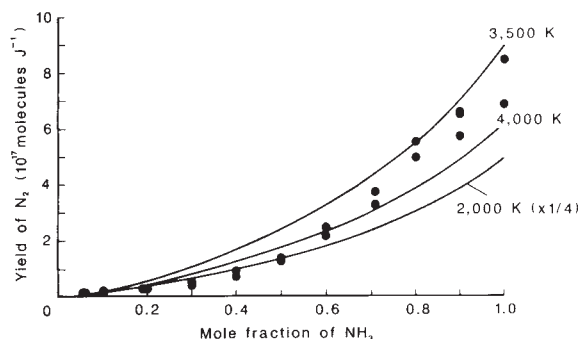


Fig. 1 Experimental production of N_2 by shock in a NH_3 - CH_4 atmosphere, filled circles. Solid lines are theoretical predictions at various freeze-out temperatures. Experimental conditions: mixtures, $NH_3 + CH_4$; pressure, 760 torr; total energy, ~ 150 J.

temperature shock in the laboratory we use the focused beam of a 1.06 μm Nd-YAG high power pulsed laser. Typically the beam contains ~ 0.2 J in a ~ 50 -ns pulse. The laser induced plasma that forms at the focus of the laser beam has been carefully studied recently with respect to its spectral properties and optical emission¹³⁻¹⁵. The initial plasma temperatures are in excess of 10,000 K but the expanding gases rapidly cool as the shock wave propagates outward from the initial volume. We feel that it provides a good laboratory simulation of the high temperature conditions associated with strong shock heating by higher velocity impacts in which gases are heated to over 2,000 K but probably does not provide an accurate simulation of the weaker shocks associated with impactors that enter with velocities less than several kilometres per second.

Gas mixtures of CH_4 - NH_3 were prepared in 250 ml glass flasks. With a lens with a 4-cm focal length, the laser beam was focused within the flask for specified duration. After irradiation, the resultant gas mixture was analysed by one of three techniques. N_2 and H_2 were measured by gas chromatography with a porapak-like column and a thermal conductivity detector. Hydrocarbons were measured by gas chromatography with a phenylisocyanate/Porasil-C column and a flame ionization detector. HCN and C_2H_2 were measured by high-resolution Fourier transform infrared spectroscopy. The two different types of analysis techniques used to determine the C_2H_2 abundance provided a check on our measurement accuracy. A further check is provided by comparing the ratio of H atoms to C plus N atoms in the products. Based on these checks and our estimates of the random and systematic errors we believe that our yields are uncertain by $\pm 30\%$. Varying amounts of water in the flasks, unavoidable in our present experimental setup¹⁶, ranging from 0.04 to 0.2% did not effect the measured yield.

Using these methods we have measured the production of N_2 in a gas mixture initially containing various amounts of NH_3

from 0.5% to 100%. These results are shown in Fig. 1. In Table 1 we list the yields for all the major gases detected for a mixture of 50% NH_3 , 50% CH_4 . In addition, propylene (C_3H_6), butane (C_4H_{10}) and pentane (C_5H_{12}) were identified in the product gases in small amounts but not quantified and at least nine other unidentified hydrocarbon products were also present in small amounts.

For comparison purposes, we have also performed theoretical calculations using a thermodynamic equilibrium model to calculate the expected yields of N_2 , HCN and hydrocarbons as described by Borucki *et al.*¹¹. Results for 'freeze-out' temperatures of 2,000, 3,500 and 4,000 K, at a pressure of 10 bar, are shown as the solid lines in Fig. 1. In these calculations graphite condensation was suppressed^{11,12}.

The experimental yields of N_2 are comparable to the theoretical yields for a freeze-out temperature of 4,000 K. The data could also, however, be fit to a lower freeze-out temperature if the theoretical yields are reduced by an efficiency factor. In particular, a freeze-out temperature of 2,000 K, which is close to the values observed for NO production by lightning in the terrestrial atmosphere⁵⁻⁷, could fit the data reasonably well with an efficiency of $\sim 25\%$. Jones and Lewis¹² have performed similar calculations and our theoretical results are in good agreement with theirs. They also use an efficiency factor in computing the final yields. The theoretical yield of N_2 drops with increasing freeze-out temperature because at higher temperatures HCN formation is favoured over N_2 . The experimental results for HCN production are fairly close to those predicted theoretically, as listed in Table 1, for a freeze-out temperature of 2,000 K. C_2H_2 , like N_2 , is about 20-30% of the theoretical value at this freeze-out temperature. Hydrogen is about half of the theoretical yield. All four of these gases are stable at high temperatures and the order-of-magnitude agreement between simple shock theory and experiments suggests that more sophisticated shock production models could explain their production. The measured H production is within 20% of that expected from the amount of C and N in the product gases ($H = 4C + 3N$).

The other hydrocarbons listed in Table 1, however, are different from their theoretical yields by large amounts and in some cases are significantly larger than the expected value. This suggests to us that shock heating is not an adequate mechanism for the production of these compounds and that other processes contribute to the yield of organic molecules besides high temperature synthesis. Photochemical reactions, driven by ultraviolet light emitted from the high temperature shock layer is a potential mechanism. The total energy radiated by a shock plasma is of the order of 1% of the total shock energy¹⁴ which, if effectively absorbed by the ambient gas, could be a possible source of the excess hydrocarbons listed in Table 1 at the observed concentrations.

Using our experimental results, we can determine if shock production of N_2 in a NH_3 - CH_4 atmosphere on the primordial Titan could account for the amount of N_2 there today. We begin by considering the last stages of accretion when the escape velocity of Titan would be highest (2.5 $km s^{-1}$) and the source region for the impacting objects would be larger, implying relative velocities at large distances, that were an appreciable fraction of Titan's orbital velocity, 6 $km s^{-1}$. Objects entering from outside the Saturn system, such as comets, would have much larger velocities but these could have been comparatively rare especially during satellite accretion. We consider satellite-forming bodies with velocities of impact in excess of 4 $km s^{-1}$. Such impactors would have been highly supersonic and generated high temperature shocks ($> 2,000$ K) as they passed through the atmosphere. Only a portion of the initial kinetic energy of such an impactor is lost to the atmosphere and of this portion $\sim 30\%$ is converted to shocks¹⁷. If we consider the accretion of the last half of Titan's mass, and assume a planetesimal size distribution that varies as r^{-3} with densities of about 2 $g cm^{-3}$ impacting into a 1 bar atmosphere, we derive a energy release

Table 1 Shock production yields for equimolar mixture of NH_3 - CH_4 .

Gas	Experimental yield (molecules J^{-1}) (method)	Theoretical yield (molecules J^{-1}) $T_f = 2,000$ K
N_2	1.2×10^{17} (a)	5.7×10^{17}
H_2	2.1×10^{18} (a)	4.0×10^{18}
HCN	2.9×10^{17} (b)	2.6×10^{17}
C_2H_2	1.1×10^{17} (b)	4.8×10^{17}
	1.5×10^{17} (c)	4.8×10^{17}
C_2H_6	8.9×10^{15} (c)	9.2×10^{13}
$CH_3C\equiv CH$	7.3×10^{15} (c)	1.4×10^{15}
C_2H_4	1.2×10^{15} (c)	2.2×10^{16}
C_3H_8	1.0×10^4 (c)	4.7×10^{13}

Detection methods: (a) thermal conductivity gas chromatography; (b) infrared spectroscopy; (c) flame ionization gas chromatography.

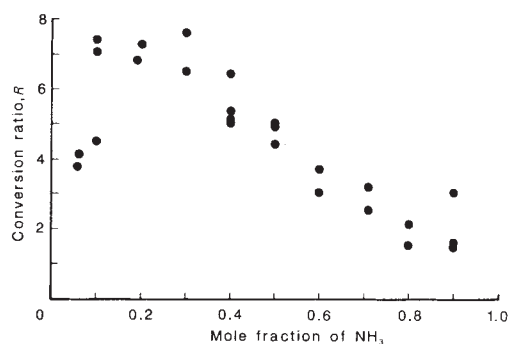


Fig. 2 Experimental ratio of moles of C converted into organic compounds per mole of N₂ produced in a shock divided by the CH₄ to NH₃ ratio in the initial mixture.

of $\sim 4 \times 10^{10} \text{ J cm}^{-2}$. We have approximated the distribution of impactors with velocities in excess of 4 km s^{-1} by considering half of the incoming mass to have exactly this velocity. As shown below, our conclusions do not depend sensitively on the exact choice of these parameter values. Shocks associated with impact ejecta would add to the energy listed above¹⁸.

If Titan formed on a timescale of 10^6 years or shorter, as seems likely¹⁹, accretional heating would have raised its surface temperature considerably higher than the present value of 95K and ammonia would have been a major constituent in the lower atmosphere. First suppose that comparable amounts of ammonia and methane were present in the lower portions of Titan's atmosphere. In that case N₂ would have been produced at an efficiency of $\sim 10^{17}$ molecules per joule of shock energy (Fig. 1). Using the shock energy of $\sim 4 \times 10^{10} \text{ J cm}^{-2}$ characterizing the later half of the satellite's formation, we find a total N₂ production of ~ 25 bar. This is more than ample to explain the current atmospheric amounts (1.5 bar) as well as the amount likely to be present in both the atmosphere and a putative ethane/methane ocean (1). If the ammonia mixing ratio in the lower portion of Titan's primordial atmosphere is reduced to 10%, then the total amount of N₂ generated would have been 2.5 bar, still enough to account for the amount present in its atmosphere and putative oceans. Allowance for production of N₂ during the satellite's early growth phase, when the escape velocity was lower than at later times, would somewhat augment the above production of N₂, while allowance for possible escape would diminish it. Finally allowance for shock chemistry of surface ices would also presumably increase the above values.

In addition to the production of N₂ the passage of meteorites through an early NH₃-CH₄ atmosphere on Titan would have produced significant organic material. The results in Table 1 suggest that for each mole of N₂ fixed, five moles of CH₄ are converted into hydrocarbons and HCN. Based on the measured H₂/N₂ ratio we can extend these results from the equimolar mixture to the other mixtures shown in Fig. 1. Figure 2 shows the number of moles of C converted into organic compounds for each mole of N₂ produced, divided by the CH₄/NH₃ in the initial mixture and the plotted as a function of the mole fraction of NH₃. We have scaled our results to the initial CH₄/NH₃ ratio so as to compensate for the effect of the starting mixture on the product yields. The falloff in the ratio at very low NH₃ mixing ratios could be due to a decline in HCN production at low initial NH₃ levels.

It is interesting to speculate on what would have been the fate of this organic material. As a result of direct condensation and further chemistry, this material would have been converted to solid organic aerosol. Such a haze would have further shielded NH₃ from solar ultraviolet light. On relative short timescales ($\sim$ years), these aerosols would have sedimented to the surface. Subsequent impacts would probably have recycled some of this material back into the atmosphere; however a kerogen-like layer may have remained as a remnant of the early processing of

Titan's atmosphere. A layer of shock produced primordial organic material could, in principle, be distinguished from organics produced photochemically over the remainder of Titan history by the isotopic and elemental ratios that may be characteristic of each processes.

We thank Larry Giver for performing the spectroscopic analysis and Joan Mathog and Jim Parris for help with the gas chromatography. This research was supported by the NASA Exobiology Program Office.

Received 13 October 1987; accepted 5 January 1988.

1. Lunine, J., Atreya, S. K. & Pollack, J. B. in *Origin and Evolution of Atmospheres* (eds Atreya, S. K., Pollack, J. B. & Matthews, M. S.) (University of Arizona Press, Tucson) (in the press).
2. Pollack, J. B. & Bodenheimer, P. in *Origin and Evolution of Atmospheres* (eds Atreya, S. K., Pollack, J. B. & Matthews, M. S.) (University of Arizona Press, Tucson) (in the press).
3. Prinn, R. G. & Fegley, B. *Astrophys. J.* **249**, 308-317 (1981).
4. Atreya, S. K., Donahue, T. M. & Kuhn, W. R. *Science*, **201**, 611-613 (1978).
5. Chameides, W. L. *Nature* **277**, 123-125 (1979).
6. Lewis, J. S., Watkins, G. H., Hartmann, H. & Prinn, R. G. in *Geological Implications of Impacts of Large Asteroids and Comets on the Earth* (eds Silver, L. T. & Schultz, P. H.) SP-190, 215-221 (Geological Society of America, Washington, 1982).
7. Park, C. & Menees, G. P. *J. geophys. Res.* **83**, 4029-4035 (1978).
8. Borucki, W. J. & Chameides, W. L. *Rev. Geophys. Space Phys.* **22**, 363-372 (1984).
9. Chameides, W. L. & Walker, J. C. G. *Origins Life* **11**, 291-302 (1981).
10. Fegley, B. Jr, Prinn, R. G., Hartman, H. & Watkins, G. H. *Nature* **319**, 305-308 (1986).
11. Borucki, W. J., McKay, C. P. & Whitten, R. C. *Icarus* **60**, 260-273 (1984).
12. Jones, T. D. & Lewis, J. S. *Icarus* **72**, 381-393 (1987).
13. Borucki, W. J., McKenzie, R. L., McLay, C. P., Duong, N. D. & Boac, D. S. *Icarus* **64**, 221-222 (1985).
14. Borucki, W. J. & McKay, C. P. *Nature* **328**, 509-510 (1987).
15. Borucki, W. J., McKay, C. P., Rogers, R. S., Boac, D. S., Duong, N. D. & Parris, J. E. *Appl. Opt.* **26**, 4319-4322 (1987).
16. Borucki, W. J., Giver, L. P., McKay, C. P., Scattergood, T. & Parris, J. E. *Icarus* (in the press).
17. Pollack, J. B., Podolak, M., Bodenheimer, P. & Christofferson, B. *Icarus* **67**, 409-443 (1986).
18. Prinn, R. G. & Fegley, B. *Earth planet. Sci. Lett.* **83**, 1-15 (1987).
19. Stevenson, D. J., Harris, A. W. & Lunine, J. I. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 39-88 (University of Arizona Press, Tucson, 1986).

Radiocarbon determination of atmospheric methane at Baring Head, New Zealand

David C. Lowe, Carl A. M. Brenninkmeijer, Martin R. Manning, Rodger Sparks & Gavin Wallace

Institute of Nuclear Sciences, Department of Scientific and Industrial Research, Lower Hutt, New Zealand

Although methane in the atmosphere is clearly increasing, its sources are still poorly defined. Measurement of carbon isotope ratios allows constraints to be placed on relative source strengths of atmospheric methane because different potential sources have different isotope ratios. Unfortunately, interpreting $^{13}\text{C}/^{12}\text{C}$ ratios of atmospheric methane is subject to uncertainty because a correction for isotope fractionation in the oxidation of methane is not well determined. Interpreting $^{14}\text{C}/^{12}\text{C}$ ratios is also complicated by the need to correct for release of $^{14}\text{CH}_4$ from nuclear power plants using rough estimates. Simultaneous use of both carbon isotope ratios, however, improves the confidence of interpretation. Here we show that measurements of the $^{14}\text{C}/^{12}\text{C}$ ratio, using accelerator mass spectrometry, and $^{13}\text{C}/^{12}\text{C}$ ratio of methane from clean Southern Hemisphere air are consistent with current estimates of both types of correction and imply that about 32% of atmospheric methane is derived from fossil carbon sources.

Recent observations of atmospheric methane in clean tropospheric air indicate that the concentration of this gas has been increasing at more than 1.5% per annum over the last decade¹⁻³. Measurements of methane in air bubbles trapped in polar ice^{4,5} suggest that the atmospheric methane concentration has been relatively constant over the past 30,000 years at less than half the present value and that the present increase started only about 400 years ago. Increasing atmospheric methane concentrations

Table 1 Carbon isotopic composition of atmospheric methane collected at Baring Head, New Zealand during southerly winds

Date collected	$\delta^{13}\text{C}$ (‰)	pMC (^{14}C) (%)	Comments
12 March 1987	-47.24 ± 0.05	102.7 ± 5.4	CO ₂ baseline
16 March 1987	-47.03 ± 0.05	106.8 ± 3.4	CO ₂ baseline
17 March 1987	-48.90 ± 0.05	102.3 ± 5.7	CO ₂ baseline
24 March 1987	-42.67 ± 0.05	112.2 ± 7.5	CO ₂ variable
10 April 1987	-46.10 ± 0.05	106.7 ± 2.6	CO ₂ baseline
16 April 1987	-48.78 ± 0.05	92.1 ± 5.7	CO ₂ baseline
1 May 1987	-41.30 ± 0.05	106.7 ± 4.1	CO ₂ variable
21 May 1987	-50.54 ± 0.05	103.2 ± 7.9	CO ₂ baseline
16 June 1987	-45.68 ± 0.05	105.4 ± 5.2	CO ₂ baseline
19 June 1987	-45.57 ± 0.05	104.0 ± 4.5	CO ₂ baseline
19 June 1987	-39.12 ± 0.05	120.5 ± 7.2	CO ₂ variable
26 June 1987	-45.96 ± 0.05	109.5 ± 3.6	CO ₂ baseline
26 June 1987	-48.59 ± 0.05	129.1 ± 6.5	CO ₂ baseline
26 June 1987	-46.37 ± 0.05	111.2 ± 5.8	CO ₂ baseline
2 July 1987	-45.80 ± 0.05	131.3 ± 5.4	CO ₂ baseline
23 July 1987	-46.62 ± 0.05	105.8 ± 10.0	CO ₂ baseline
23 July 1987	-38.33 ± 0.05	125.0 ± 3.6	CO ₂ variable
29 July 1987	-43.69 ± 0.05	115.4 ± 3.4	CO ₂ variable

Methane data are grouped according to a 'steadiness' criterion applied to concurrent atmospheric carbon dioxide measurements made at the Baring Head site and are similar to those described in refs 15 and 16. For the purposes of the methane isotopic measurements reported here two broad classifications are made: (1) 'baseline', where the atmospheric CO₂ data showed a uniform value indicative of well mixed background air; and (2) variable, where the atmospheric CO₂ measurements indicate variability caused by local sources and/or sinks of CO₂. The methane ^{13}C data are listed as parts per mil (‰) with respect to PDB as determined by conventional mass spectrometry. The ^{14}C data are given as % modern carbon (pMC) as determined by accelerator mass spectrometry.

are of concern because the gas is a strong absorber in the infrared, and if current rates of increase continue it will contribute significantly to global warming through the 'greenhouse effect'^{6,7}. In addition, methane regulates tropospheric OH radical concentrations and is a net source of ozone, hydrogen and carbon monoxide⁸.

Although progress has been made in identifying sources of atmospheric methane and quantifying their fluxes⁹⁻¹¹, there are still major uncertainties in methane budget estimates. Stevens and Rust¹² have suggested that measurements of $^{13}\text{C}/^{12}\text{C}$ in methane in clean air may help to determine the major sources of gas. But both they and Tyler¹³ have shown that methane in clear air is enriched in ^{13}C by about 10–25‰ compared with its major biogenic sources. Some of this enrichment can be explained by the kinetic isotope effect in the oxidation of methane by the OH radical. Methane containing the heavier isotope has a smaller rate coefficient for the oxidation reaction and hence a longer lifetime in the atmosphere. The corresponding enrichment has recently been reported¹⁴ as $10 \pm 7\%$. But the relatively large uncertainty in this value and the spread in $^{13}\text{C}/^{12}\text{C}$ ratios for various methane sources limits the use of ^{13}C for resolving the uncertainties in the global methane budget.

To evaluate the relative strengths of biogenic and fossil sources of atmospheric methane we have begun a series of radiocarbon determinations of carbon in methane extracted from clean air. Because atmospheric methane concentrations in clean air are low, about 1.6 parts per million (p.p.m.) in the Southern Hemisphere, we have used accelerator mass spectrometry to make the radiocarbon analyses. All the measurements reported here are for samples collected at a baseline CO₂ monitoring station at Baring Head, New Zealand. This station is at 41°S but wind trajectory analyses¹³ show that, during southerly winds, air sampled at this site can be taken as representative of the lower troposphere at around 55°S. Monitoring records over 15 years^{15,16} show that the short term steadiness of the continuously

measured atmospheric CO₂ concentration is a reliable indicator of baseline conditions.

Measurements by Fraser *et al.*³ at Cape Grim, Tasmania, at the same latitude as Baring Head, demonstrate that under baseline conditions methane concentrations show very little variation about a steady seasonal cycle with a small secular increase. A baseline methane concentration of 1.63 p.p.m. was found at Cape Grim in July 1987 (P. J. Fraser, personal communication). Three flask samples taken from Baring Head in that month during baseline conditions gave methane concentrations measured by Fraser ranging from 1.628 to 1.634 p.p.m. Thus, although methane concentration measurements are not available for all the samples reported here, we are confident that for the baseline samples the concentration will, as expected for Cape Grim, vary from 1.60 to 1.63 p.p.m. over the sampling period.

A complete description of the methane sampling and analysis methods used in this study will be given elsewhere but a brief description follows. Methane is extracted from ~1 m³ of air using a cryogenic technique. During sampling, air is passed through a series of traps which selectively remove carbon dioxide, water vapour and non-methane hydrocarbons, leaving methane and a mixture of rare gases trapped with O₂ and some N₂ in an activated charcoal matrix at liquid nitrogen temperature. Samples are returned to the laboratory where, after removal of carbon monoxide, the methane is burnt in oxygen to carbon dioxide using a platinized catalyst at 750 °C. The $^{13}\text{C}/^{12}\text{C}$ ratio of the CO₂ derived from the methane is measured by conventional high precision mass spectrometry, and the same CO₂ gas sample is then used in the preparation of graphite targets¹⁸ for the measurement of $^{14}\text{C}/^{12}\text{C}$ ratio by accelerator mass spectrometry. The graphite targets contain from 0.5 to 0.8 mg of carbon depending on the collection efficiency of the sampling apparatus and the volume of air sampled.

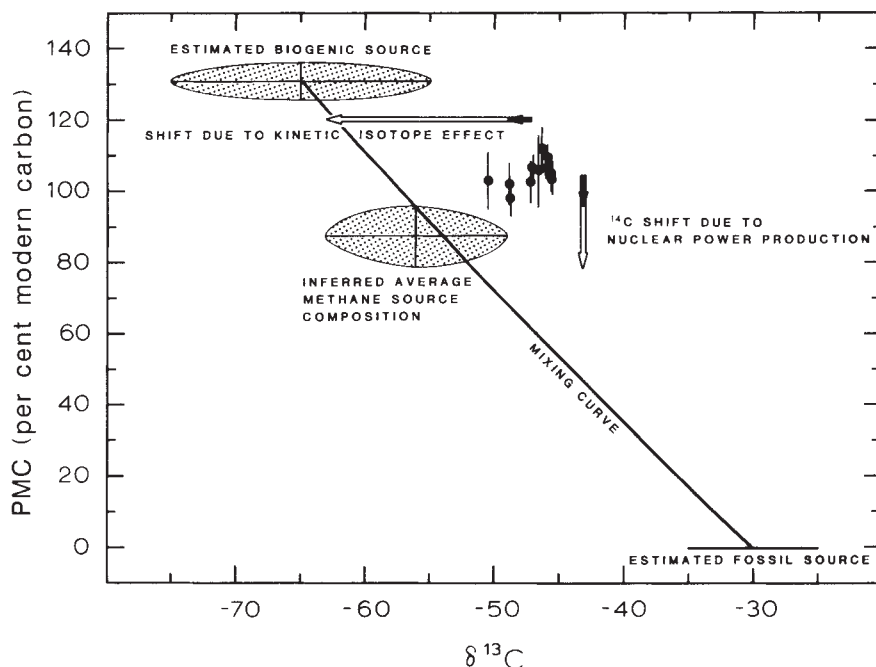
The results for 18 samples are shown in Table 1 using the standard definitions¹⁷ of $\delta^{13}\text{C}_{\text{PDB}}$ and per cent modern carbon (pMC). Note that pMC values are obtained from sample $^{14}\text{C}/^{12}\text{C}$ ratios by applying a fractionation correction to normalize all values to a $\delta^{13}\text{C}$ of -25% and consequently pMC values are not affected by the kinetic isotope effect. All samples were taken during southerly wind conditions but are classified as being baseline samples only when CO₂ concentrations measured at the same time meet steadiness criteria defined by Manning and Pohl¹⁵. Among the samples taken under baseline conditions, three (16 April, 26 June and 23 July 1987) appear to be inconsistent with the remainder. Excluding these leaves ten measurements for which the mean $\delta^{13}\text{C}$ is $-47.0 \pm 1.6(1\sigma)\%$, the corresponding mean pMC being $106 \pm 3\%$. The results for variable southerly conditions indicate heavier methane with a $\delta^{13}\text{C}$ of $-41.0 \pm 2.3\%$, and pMC of $116 \pm 7\%$. The $\delta^{13}\text{C}$ value for baseline conditions is in close agreement with values measured in clean air in the Northern Hemisphere¹³.

To see how these results constrain the relative fractions of fossil and biogenic methane, we consider a two-source model. One source is taken to be the average of all fossil methane sources and the other to be the average of all biogenic methane sources.

Measured $\delta^{13}\text{C}$ values for fossil fuels range from -35 to -25% , while geothermal and volcanic methane show variable $\delta^{13}\text{C}$ values¹⁹. A representative value for the fossil source will be taken as $-30 \pm 5\%$. The fossil methane source contains no ^{14}C , corresponding to a pMC of zero. Measured $\delta^{13}\text{C}$ values of biogenic methane sources range from -75 to -55% ¹³ and a representative value of $-65 \pm 10\%$ will be used here for the average biogenic source. However, there are no direct measurements of $^{14}\text{C}/^{12}\text{C}$ ratios for biogenic sources so these must be inferred.

Methane released from animal, agricultural and similar sources can be expected to have a pMC value close to that of contemporary atmospheric CO₂. However, other biogenic sources,

Fig. 1 The carbon isotope ratios of atmospheric methane sampled under baseline conditions at Baring Head, New Zealand, are shown here with the data for a two-source model used to estimate the fraction of atmospheric methane derived from fossil carbon. Selected baseline data are shown as full circles with error bars. The horizontal arrow represents the shift in $\delta^{13}\text{C}$, $10 \pm 7\%$, determined by Davidson *et al.*¹⁴ for the kinetic isotope effect in the oxidation of atmospheric methane by the OH radical. The vertical arrow represents the shift in pMC estimated to be due to emissions of $^{14}\text{CH}_4$ by the nuclear power industry. Combining these two shifts and their uncertainties with the isotope ratios observed for atmospheric methane at Baring Head, leads to an inferred average source in the shaded ellipse at the centre of the figure. This lies on the mixing curve between the biogenic and fossil sources and its position implies that the current fossil fraction of atmospheric methane is 32%.



such as biomass burning and wetlands, release carbon that has been stored for some time after photosynthesis from the atmosphere. Furthermore, as atmospheric methane has a non-zero lifetime, what is observed in the atmosphere now is the result of release over many years past. These factors must be taken into account together with the record of ^{14}C from nuclear weapons testing, in order to determine a pMC value for the average biogenic source. We have taken the global atmospheric $^{14}\text{CO}_2$ record as the average of long-term data records reported by Levin *et al.*²⁰ for Vermunt, Austria, and unpublished ^{14}C data from Makara, New Zealand. Taking the methane mean lifetime in the atmosphere as 7.6 years, and biosphere residence times of 0 or 5 years, gives estimated pMC values of 128.2 and 134.9% respectively for methane of biogenic origin. The average biogenic source will be taken to have a pMC value of $131 \pm 3\%$.

The $\delta^{13}\text{C}$ and pMC values measured for clean-air methane differ from the result of mixing the biogenic and fossil sources due to two further factors. The first is the kinetic isotope effect mentioned above. If the methane concentration in the atmosphere was constant this would cause the observed $\delta^{13}\text{C}$ value to be higher by $10 \pm 7\%$. As the atmospheric concentration is increasing, however, there is a small displacement from this equilibrium. Assuming the increase over the last few decades has been exponential at the observed rate of 1.5% per annum the shift in $\delta^{13}\text{C}$ between the total methane source and the value in clean air will be $9 \pm 7\%$.

The second factor to be taken into account is an increase in the $^{14}\text{C}/^{12}\text{C}$ ratio due to release of $^{14}\text{CH}_4$ from nuclear power production. I. Levin (personal communication) has made estimates of this $^{14}\text{CH}_4$ release, which is primarily from pressurized water reactors and is expected to be increasing in proportion to nuclear power production from this type of reactor. The total release of $^{14}\text{CH}_4$ still in the atmosphere in 1987 is estimated as 141 TBq (3,800 Ci); however, the amount in the Northern Hemisphere will exceed that in the Southern Hemisphere as the vast majority of sources are north of the equator. Using interhemispheric exchange times of 1 or 2 years with Levin's production rates gives effective shifts of pMC values for the Southern Hemisphere of 18.7 or 17.8% respectively. However, as there is still much uncertainty in the production rate estimates we will take the correction for nuclear power plant methane as $18.3 \pm 9.0\%$.

These factors are summarized in Fig. 1. It can be seen that the corrections for the kinetic isotope effect and nuclear power production bring the observed carbon isotope ratios within the

stated uncertainties of the mixing curve joining the two average sources. The most likely fossil methane fraction based on these figures is 32%. Adjusting the mixing line to minimize the fossil fuel fraction while remaining within the uncertainties given, would give a minimum value of 23%. With reference to Fig. 1, it should be noted that the ability of ^{13}C measurements to define relative source strengths is limited by uncertainty in the kinetic isotope effect. On the other hand, the ability of ^{14}C measurements to define relative source strengths is limited by uncertainty in the nuclear power production correction.

A continuing programme of isotope measurements is needed to track variations in relative source strength of atmospheric methane. As methane has a mean life in the atmosphere of about 7 years, the fossil methane fractions given above refer to source emissions averaged over this time scale. Wahlen (personal communication) has reported measurements of $^{14}\text{C}/^{12}\text{C}$ ratios in methane in the Northern Hemisphere which show higher pMC values than reported here. While the nuclear power production of $^{14}\text{CH}_4$ is consistent with higher values in the Northern Hemisphere, a more detailed comparison of results from both hemispheres extending over several years is needed to resolve this source clearly. Unfortunately the ability of ^{14}C measurements to determine the fossil methane component of atmospheric methane is progressively diminished as nuclear power production of $^{14}\text{CH}_4$ increases.

We thank Paul Fraser for carrying out methane concentration measurements from flask samples taken at Baring Head, Ingeborg Levin for supplying details of her estimates of $^{14}\text{CH}_4$ released by nuclear power plants, and Dave Bibby for initial advice on methane extraction techniques.

Received 12 October 1987; accepted 8 February 1988.

1. Khalil, M. A. K. & Rasmussen, R. A. *J. geophys. Res.* **88**, 5131-5144 (1983).
2. Blake, D. R. & Rowland, F. S. *J. atmos. Chem.* **4**, 43-62 (1986).
3. Fraser, P. J., Hyson, P., Rasmussen, R. A., Crawford, A. J. & Khalil, M. A. K. *J. atmos. Chem.* **4**, 3-42 (1986).
4. Craig, H. & Chou, C. C. *Geophys. Res. Lett.* **9**, 1221-1224 (1982).
5. Pearman, G. I., Etheridge, D., de Silva, F. & Fraser, P. J. *Nature* **320**, 248-250 (1986).
6. Laci, A., Hansen, J., Lee, P., Mitchell, T. & Lebedeff, S. *Geophys. Res. Lett.* **8**, 1035-1038 (1981).
7. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547-5566 (1985).
8. Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210-7254 (1981).
9. Bingemer, H. G. & Crutzen, P. J. *J. geophys. Res.* **92**, 2181-2187 (1987).
10. Holzapfel-Pschorn, A. & Seiler, W. *J. geophys. Res.* **91**, 11803-11814 (1986).
11. Crutzen, P. J., Aselmann, I. & Seiler, W. *Tellus* **38B**, 271-284, (1986).
12. Stevens, C. M. & Rust, F. E. *J. geophys. Res.* **87**, 4879-4882 (1982).
13. Tyler, S. C. *J. geophys. Res.* **91**, 13232-13238 (1986).

14. Davidson, J. A., Cantrell, C. A., Tyler, S. C., Shetter, R. E., Cicerone R. J. & Galvert, J. G. *J. geophys. Res.* **92**, 2195–2199 (1987).
15. Manning, M. R. & Pohl, K. P. INS-R-350. *Institute of Nuclear Sciences* (DSIR, Lower Hutt, New Zealand, 1986).
16. Lowe, D. C., Guenther, P. R. & Keeling, C. D. *Tellus* **31**, 58–67 (1979).
17. Stuiver, M. & Polach, H. A. *Radiocarbon* **19**, 355–363 (1977).
18. Lowe, D. C. & Judd, W. J. *Nucl. Instrum. Meth. phys. Res. B* **28**, 113–116 (1987).
19. Lyon, G. L. & Hulston, J. R. *Geochim. cosmochim. acta* **48**, 1161–1171 (1984).
20. Levin, I. et al. *Radiocarbon* **27**, 1–19 (1985).

High oxide ion conductivity in $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$

M. Lacerda*, J. T. S. Irvine, F. P. Glasser & A. R. West

University of Aberdeen, Department of Chemistry, Meston Walk, Aberdeen AB9 2UE, UK

Oxide ion conducting solids have much potential for use in fuel cells, sensors, oxygen pumps and gauges. Existing oxide ion conductors are essentially confined to a small group of ceramic oxides with structures related to fluorite, the most important examples being ZrO_2 - and Bi_2O_3 -related materials. There is a clear need for new oxide ion conductors with improved properties. We report measurements of the conductivity of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$, which indicate it to be an oxide ion conductor with bulk conductivity only 8–10 times less than that of yttria-stabilized zirconia. The structure of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ is unrelated to the fluorite structure. The reasons for the high conductivity are not well understood but may be associated with the fact that the structure contains a mixture of anions, a three-dimensional aluminate framework and essentially free oxide ions.

The aluminate $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ is a stable ceramic material, well known from its occurrence in cement clinkers. It can be prepared by solid state reaction of CaCO_3 and Al_2O_3 mixtures in platinum crucibles at 1,200–1,350 °C for 1–2 days; its melting point is variously quoted as being in the range 1,360–1,400 °C¹.

The structure of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ is cubic and may be described as a three-dimensional framework formed by the linking of eight-membered rings of AlO_4 tetrahedra which link up by sharing either three or four of their corners with other tetrahedra. However, only 32 of the 33 oxygens in the formula unit belong to the aluminate framework. The additional oxide ions and the calcium ions occupy sites within the aluminate framework. Questions remain over the exact stoichiometry of the material. If the material is free from hydrogen-containing species or other anions, its stoichiometry can be written as $\text{Ca}_{12}(\text{Al}_{14}\text{O}_{32})\text{O}$. When prepared in air, however, it appears to contain some superoxide ions² and is water-sensitive, even at high temperatures, presumably due to the formation of hydroxide ions. It can also be prepared with a significant proportion of fluoride or chloride ions, up to the limiting composition $\text{Ca}_{12}\text{Al}_{14}\text{O}_{32}(\text{F}, \text{Cl})_2$ (ref. 3). The structure of the fully fluorinated compound has been determined⁴.

For conductivity measurements, a pressed pellet was sintered in air by firing at 1,340–1,360 °C; platinum paste was painted on to opposite pellet surfaces and hardened by firing at 900 °C. The pellet was then attached to the platinum leads of a conductivity cell, which was placed in laboratory air inside a tube furnace whose temperature was controlled and measured to within 2 °C. Conductivities were measured in the form of a.c. impedance using a Solartron 1250 Frequency Response Analyser linked to a 1286 Electrochemical Interface. Data were plotted in the complex impedance plane and analysed to achieve separation of bulk, grain boundary and electrode effects. The prime objectives were to determine the bulk crystal conductivity as a function of temperature and atmosphere and to find appropriate conditions for reducing the magnitude of the grain boundary resistance in the samples.

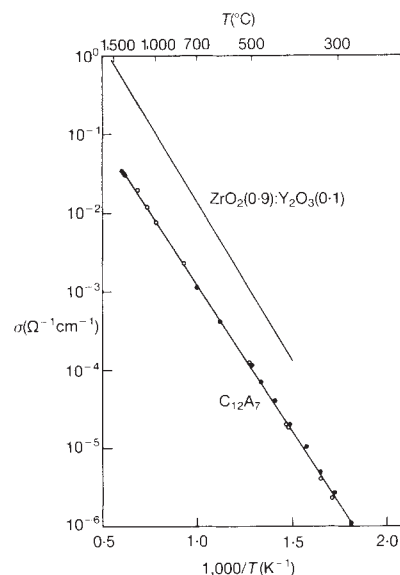


Fig. 1 Arrhenius plots of the conductivity of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ (C_{12}A_7), compared with yttria-stabilized zirconia⁵. ●, heating cycles. ○, cooling cycles.

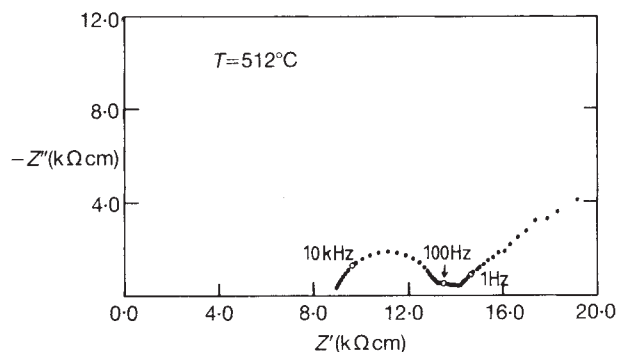


Fig. 2 Typical a.c. impedance plot for $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$.

In Fig. 1, the bulk conductivity is shown over the temperature range 280–1,350 °C. Data are shown on both heating and cooling cycles and for more than one sample. The data are remarkably reversible and reproducible. The plot is linear with an activation energy of 0.74 eV and a prefactor of 0.81. Data for yttria-stabilized zirconia⁵ are shown for comparison; it can be seen that the conductivity of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ is less than that of zirconia by a factor of 8–10, depending on temperature.

The bulk conductivity of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ is sensitive to sample history. Its value is optimized by heating at temperatures of ~1,350 °C, close to the melting point. Annealing at lower temperatures, 900–1,300 °C, reduces the conductivity, probably because of water uptake by the sample which occurs spontaneously, despite the high temperatures⁶. Once prepared at 1,350 °C, however, the samples appear to be stable in air at temperatures below ~700 °C.

The a.c. impedance results showed that a grain boundary resistance was also usually present, but could be reduced by firing at ~1,350 °C. Impedance data in Fig. 2 show a semicircle with a frequency maximum at ~4 kHz and an associated capacitance, given by the relation, $2\pi fRC = 1$ at the semicircle maximum, of 20 nF. From the magnitude of its capacitance, this semicircle is attributed to a grain boundary effect, of resistance given by the diameter of the semicircle. At this particular temperature, the grain boundary resistance is about 4 kΩ, which is considerably less than the bulk resistance, given by the high frequency intercept of the data on the real Z' axis, of about 9 kΩ. At lower temperatures the high frequency semicircle,

* Permanent address: Universidade Nova de Lisboa, Department of Physics, 2825 Monte da Caparica, Lisboa, Portugal.

corresponding to the bulk resistance, was seen with an associated capacitance of 2 pF. At frequencies below ~ 100 Hz in Fig. 2, a small poorly resolved semicircle with a resistance of 2 k Ω and an inclined spike are seen. Both of these are attributed to electrode effects. The semicircle, with associated capacitance of 6 μ F, is attributed to transfer of oxygen/oxide across the pellet/electrode interface and the inclined spike to a Warburg effect associated with diffusion-limited migration of oxygen away from the electrode-pellet interface.

To confirm that the conducting species were oxide ions, a concentration cell was made such that opposite pellet surfaces could be exposed to different gases. The open circuit voltage was measured at one atmosphere total pressure with pure oxygen in one gas compartment and air in the other. From the Nernst equation, $E = (RT/nF) \ln(p''_{O_2}/p'_{O_2})$, and if the transport number of oxide ions is unity, a voltage of 50 mV at 1,250 °C is expected, assuming $n = 4$ per mole oxygen gas. The experimental value was 45 mV, close to the theoretical value, thereby confirming that the current carriers were ionic and not electronic. If the conducting species were either peroxide or superoxide ions, lower e.m.f. values, given by $n = 2$ and 1 respectively, would have been expected. These results do not rule out the possibility that hydrogen-containing species, such as OH $^-$, are the current carriers. However, the conductivity was reduced markedly by water uptake, and this may be explained by the replacement of mobile oxide ions by hydroxide ions, according to the chemical reaction:



Further studies are underway to understand the conduction mechanism and to optimize the conductivity of both bulk and grain boundary regions.

Received 30 December 1987; accepted 5 February 1988.

1. *Phase Diagrams for Ceramists* No. 4308 (American Ceramic Society, 1975).
2. Imlach, J. A., Dent Glasser, L. S. & Glasser, F. P. *Cement & Concrete Res.* **1**, 57-61 (1971).
3. Jeevaratnam, J., Glasser, F. P. & Dent Glasser, L. S. *J. Am. Ceram. Soc.* **47**, 105-106 (1964).
4. Williams, P. P. *Acta crystallogr.* **B29**, 1550-1551 (1973).
5. de Pous, O. *World Ceramics* **2**, 99-111 (1985).
6. Singh, V. K. & Glasser, F. P. *Ceramics Int.* (in the press).

Stable perinaphthenyl radicals in flints

Harish Chandra*, Martyn C. R. Symons*
& Dafydd R. Griffiths†

* Chemistry Department, Leicester University,
Leicester LE1 7RH, UK

† Institute of Archaeology, University College,
London WC1H 0PY, UK

The technique of electron spin resonance (ESR) spectroscopy is proving useful in determining dates by the study of radical species in hard geological and archaeological samples. The key aspect of this work is that normally unstable radical centres, when formed in rigid solids, are rendered inactive by their immobility. Such species may be formed by ionizing radiation, or by thermal treatment. Flint, being an extremely hard form of polycrystalline silica, provides an ideal trapping medium, and we have recently established that methyl radicals (CH $_3$), formed thermally in various flint samples from unknown precursors, are trapped in cavities and are stable up to at least 350 °C. We have now found that another organic radical, perinaphthenyl (or phenalenyl, compound I in Fig. 1) is also formed in a variety of flint samples. ESR spectra provide an unambiguous detection of these radicals, and show that they are rotating rapidly at room temperature. Dimers are not observed, indicating that they are individually isolated in flint cavities.

Three types of radicals have been detected by ESR spectroscopy

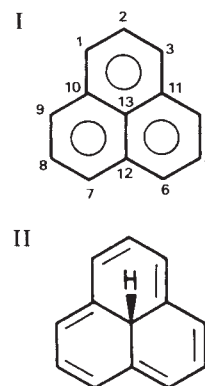


Fig. 1 Structure of the perinaphthenyl radical (I) and a possible precursor (II).

copy in a range of flint samples¹⁻⁴. One type, which may be termed intrinsic, is based on the silica itself, and is formed by ionizing radiation. These radicals are essentially electron-capture and electron-loss centres, which in principle can be used as part of a dating routine as the signals are proportional to the radiation dose^{3,4}.

A second important centre has been assigned to carbon particles formed on heating from organic matter occluded in the flint. This single broad line is only present in flints heated above a threshold temperature of ~ 400 °C. It is thermally quite stable, and can be used to assess the sample's history of heat treatment used to improve the mechanical properties of flint artefacts².

The third and most surprising type of centre is due to organic radical species such as methyl radicals. These are formed thermally rather than radiolytically. From the form of the spectra, we deduce that they are trapped singly in cavities in silica that are large enough to permit free rotation and small enough to prevent migration. They are formed by pyrolysis, but their source is a matter of speculation¹.

In some samples, a multi-line spectrum grew in on heating, quite independently of the methyl radical spectrum (Fig. 2). This has now been identified as the perinaphthenyl radical (I). This radical, originally studied by Sogo *et al.*⁵ using ESR spectroscopy, is remarkably stable. It is the simplest possible neutral π -radical species involving fused benzene rings, the unpaired electron being delocalized symmetrically through the π -system, as shown by the fact that there are six-equivalent strongly coupled protons (at positions 1, 3, 4, 6, 7 and 9; $A(^1\text{H}) = 6.3$ G) and three weakly coupled protons (at positions 2, 5, 8; $A(^1\text{H}) = 1.8$ G (ref. 6)). In fluid solution (as opposed to the situation in flint), the radical readily dimerizes, the ESR signal being lost reversibly in the -15 to -20 °C region⁷.

Our results, including the g -value of 2.00268, are in good agreement with published data and identification of the radical is virtually unambiguous. The lines are isotropic, and extremely narrow for a 'solid-state' spectrum (< 0.2 G between points of maximum slope), despite the fact that considerable anisotropy would be expected. This probably means that the radicals are completely free to rotate rapidly about all axes, which implies large spherical cavities. It is possible for rapid rotation to be restricted to the z -axis (normal to the aromatic plane) because for α -protons (protons bound directly to the radical centre), A_z is ~ 0.5 ($A_x + A_y$). This would require disc-shaped cavities that would be less likely than nearly spherical cavities in a silica matrix. The fact that the intensity is invariant in the 0-100 °C range means that dimerization is completely inhibited by the cages. Also, the absence of spin-flip satellite lines at high microwave powers^{8,9} strongly suggests that the immediate environment of each radical does not contain protons. We conclude that each cavity originally contained only one radical precursor and that formation does not occur in regions containing many

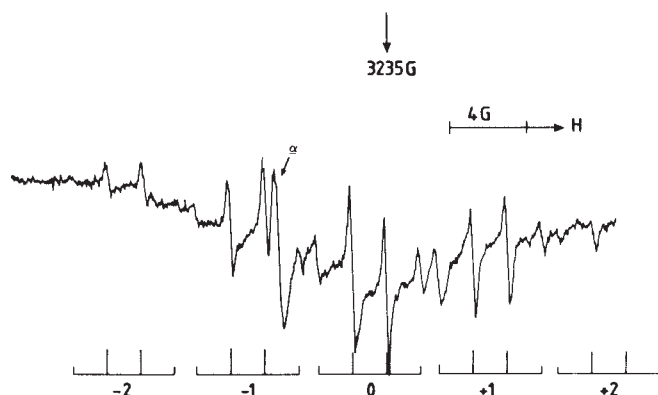


Fig. 2 First derivative ESR spectrum for a flint taken from Grimes Graves, after heating to 400 °C. The $M_I(^1\text{H}) = \pm 3$ set of quartets are not shown, but were detectable at high gain and high modulation. The line marked α ($g = 2.0054$) is possibly due to carbon. The broad underlying resonance, which was well defined at 77 K and high microwave powers, is due largely to an aluminium centre induced by ionizing radiation⁴.

hydrocarbon molecules. Thus these radicals are probably not formed in the same organic regions that are responsible for the polymeric carbon signals.²

There may be several sources of these radicals. It has been shown that pyrolysis of a wide range of hydrocarbons in the 450° to 750 °C range gives low yields of perinaphthenyl radicals⁶. Our thermal results are not in accord with this as the radical signals appear at ~275 °C, reach a maximum in the 300–400 °C range, and are lost above ~500 °C. In our case, the precursors must be relatively unstable and the most obvious candidate is compound (II) shown in Fig. 1. This, however, is a relatively rare compound and it is unlikely that it is present in sufficient concentration during the growth of a range of different flints. The task of identifying the precursor(s) is rendered extremely difficult because of the low maximum concentrations of perinaphthenyl radicals, and the fact that most occluded organic compounds are unlikely to form these radicals on heating in this range. Thus the precursors are probably only minor constituents of the total of occluded compounds.

The fact that the perinaphthenyl radicals we observe are lost at temperatures well below those used for generation in ref. 6 seems surprising, but can probably be explained by significant differences between the experiments. These include: (1) thermal contact time, which was short in Bennett's experiments but extensive in ours; (2) the fact that generation and loss compete in the hydrocarbon pyrolysis, whereas our results suggest that formation is complete before the onset of decomposition; and (3) that each radical is isolated in cavities, our decompositions are either unimolecular, or involve the silica cage in some way, whereas the gas-phase reactions probably involve bi- or ter-molecular processes.

We have found no pattern in the types of flints which give the perinaphthenyl radical (I) on heating. They include several flints from Grimes Graves and a range of other British and middle-eastern cherts (for example, Ksarakil flints). Several archaeologically heated samples gave this signal on measurement at room temperature without laboratory heating. In those cases it is possible that a plot of signal intensity versus the temperature of reheating would enable us to estimate the maximum temperature to which the sample had been heated in the past. This is not, however, a generally applicable approach as many flints show no sign of this species.

It is clear that ESR spectroscopy is a powerful tool for determining dates of geological and archaeological samples. Because of the extremely hard and close-knit structure of flints, the damage (radical) centres caused by ionizing radiation are permanently trapped. The ESR signal that results is then propor-

tional to the number of centres, which is proportional to the dose of ionizing radiation. This, in turn, is a function of the age of the sample. When a flint is heated, these centres are lost, so the 'clock' is reset. Fortunately, our forebearers were in the habit of heating flint, both in their use as hearth stones, and to facilitate working them into sharp instruments. Analysis of the intrinsic radical centres by ESR enables such activity to be dated. Another use of ESR is to establish whether or not a flint has been heated. Such heating causes occluded organic material to char, and the resulting carbon also contains trapped radicals, which give a well-defined ESR signal.

The fact that perinaphthenyl radicals are sometimes thermally generated may help us to obtain a fine tuning of the extent of heating. Alternatively, their present may be linked to the source of the flint, as only a small number of samples from different sources give the signal. So far, the quantities studied have not been sufficient to give us any clues as to the precursor of these unexpected radicals.

We thank Dr Brian Tabner and Dr N. J. Seeley for helpful discussion.

Received 3 November 1987; accepted 26 January 1988.

1. Griffiths, D. R. *et al. Nature* **300**, 435–437 (1982).
2. Robins, G. V., Seeley, N. J., Symons, M. C. R. & McNeil, D. A. C. *Nature* **276**, 703–704 (1978).
3. Robins, G. V., Seeley, N. J., Symons, M. C. R. & McNeil, D. A. C. *Nature* **300**, 435–436 (1982).
4. Griffiths, D. R., Seeley, N. J., Chandra, H. & Symons, M. C. R. *PACT* **9**, 399–410 (1983).
5. Sogo, P. B., Nabazaki, M. & Calvin, M. *J. chem. Phys.* **26**, 1343–1352 (1957).
6. Bennett, J. B. *Proc. chem. Soc.* **144**–145 (1961).
7. Gerson, F. *Helv. chim. Acta* **49**, 1463–1472 (1966).
8. Zeldes, H., Trammell, G. T. & Livingston, R. J. *chem. Phys.* **32**, 618–625 (1960).
9. Symons, M. C. R. *Chemical and Biochemical Aspects of Electron Spin Resonance Spectroscopy* (Van Nostrand Reinhold, Wokingham, UK, 1978).

Cobalt and copper distributions in the waters of Santa Monica Basin, California

Kenneth S. Johnson, Paul M. Stout*, William M. Berelson† & Carole M. Sakamoto-Arnold*

Moss Landing Marine Laboratories, PO Box 450, Moss Landing, California 95039-0450, USA

* Marine Science Institute, University of California, Santa Barbara, California 93106, USA

† Department of Geological Sciences, University of Southern California, Los Angeles, California 90089-0741, USA

The trace metals cobalt and copper are removed from the oceans interior by scavenging on to particle surfaces¹, but the mechanisms for removal of these two metals are probably quite different. Cobalt appears to be scavenged by manganese oxide particles^{1,2}, whereas organic compounds are the main carrier phase for copper^{3,4}. Remobilization of these metals in marine sediments therefore proceeds by different pathways. The differences in the pathways of remobilization are accentuated in oxygen-deficient environments: manganese oxide reduction is accelerated at low oxygen levels⁵ and organic carbon is preserved⁶. Cobalt fluxes from sediments underlying oxygen-deficient waters should be enhanced and copper fluxes reduced. We report here measurements of the cobalt and copper distributions in the waters of an oxygen-deficient marine basin in the Southern California Bight. Cobalt concentrations near the bottom are raised four times above the background level, whereas copper concentrations show no increase. These measurements confirm features of existing models for the oceanic cycles of these metals.

A strong oxygen minimum (<10 μM O₂) is present in the Southern California Bight from 400 to 900 m. The sill depth of

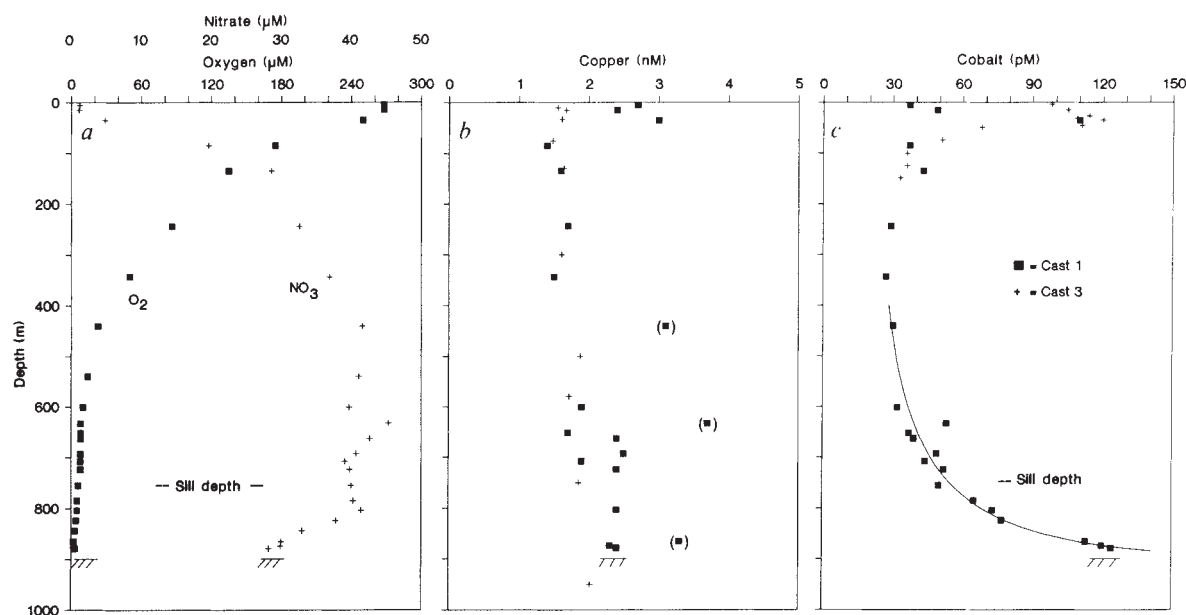


Fig. 1 Vertical profiles of *a*, nitrate and oxygen concentrations, *b*, copper concentrations and *c*, cobalt concentrations, measured in the water column of Santa Monica Basin on hydrocasts 1 and 2. Copper concentrations measured at a station off the California coast¹⁴ (+) are shown for comparison with the values we measured (■). The solid line through the cobalt data, $\log_e \text{Co (pM)} = 5.516 - 0.1561 \times (900 - Z)^{0.5} + 0.00264 \times (900 - Z)$ was fitted by least squares.

Table 1 Trace metal concentrations and associated hydrographic properties measured in Santa Monica Basin

Depth (m)	<i>T</i> (°C)	Salinity	O ₂ (μM)	NO ₃ (μM)	Co (pM)	Cu (nM)
Cast 1: 17 March 1987, 33°47'83" N, 118°53'05" W						
5	14.98	33.410	268	1.2	37	2.7
15	—	33.416	268	1.1	49	2.4
35	—	33.468	250	4.8	110	3.0
85	10.56	33.649	175	19.6	37	1.4
135	9.31	33.937	135	28.6	43	1.6
244	8.40	34.144	86	32.6	29	1.7
343	7.50	34.205	50	36.9	27	1.5
440	6.76	34.268	23	41.6	30	(3.1)
540	6.20	34.329	14	41.1	(292)	(12.0)
601	5.77	34.351	10	39.7	32	1.9
652	5.50	34.370	8	—	37	1.7
708	5.30	34.379	8	39.1	44	1.9
Cast 2: 17 March 1987, 33°48'16" N, 118°52'21" W						
633	5.59	34.365	8	45.3	53	(3.7)
663	—	34.372	8	42.6	39	2.4
693	5.44	34.374	8	40.7	49	2.5
724	5.31	34.377	8	39.8	52	2.4
755	5.27	34.383	6	40.0	50	—
785	5.21	34.387	5	40.3	65	—
804	5.16	34.388	5	41.4	73	2.4
824	5.16	34.388	4	37.8	77	—
844	5.13	34.391	3	33.0	—	—
865	5.11	34.392	2	30.0	113	(3.3)
874	5.10	34.396	2	29.9	120	2.3
879	5.11	34.394	3	28.2	124	2.4
Cast 3: 22 March 1987, 33°44'28" N, 118°45'69" W						
4	12.23	33.549	248	8.8	98	—
15	—	33.552	245	8.8	105	—
26	12.17	33.553	243	9.0	114	—
31	12.20	33.552	248	9.1	109	—
35	12.18	33.555	239	9.3	120	—
45	12.07	33.553	234	9.5	111	—
49	11.41	33.603	186	13.5	68	—
74	—	33.725	158	15.7	51	—
100	9.82	33.842	142	21.4	36	—
125	9.48	33.896	137	22.8	36	—
149	9.20	33.998	122	24.4	33	—

Concentrations enclosed by parentheses are believed to have been contaminated during the sampling process, being reproducible in the laboratory but apparently oceanographically inconsistent. Missing data are indicated by a dash.

the Santa Monica Basin (745 m), which lies offshore of Los Angeles, California, is coincident with this minimum. Oxygen concentrations decrease continuously within the basin (Fig. 1a) from the already low values ($8 \mu\text{M O}_2$) at the sill depth to values of $<1 \mu\text{M O}_2$ at the bottom (900 m). Nitrate reduction within the sediments causes a flux of nitrate into the sediments and an accompanying decrease in nitrate concentrations towards the bottom (Fig. 1a)⁷. Extremely low oxygen concentrations prevent colonization of the basin by macrofauna and the sediments are laminated^{8,9}. Forty to seventy per cent of the organic carbon reaching the sediments is preserved in this and other oxygen-deficient basins of the Southern California Bight^{6,7}. In comparison, $<10\%$ of the organic carbon is preserved in pelagic sediments⁶. Particulate manganese oxides reaching the sea floor would be expected to be consumed very near the sediment-water interface in this basin as a consequence of the very low oxygen concentrations.

Water samples were collected from the surface to the bottom of the water column in Santa Monica Basin using standard five-litre Niskin bottles hung on a stainless steel hydrowire. Cobalt and copper concentrations were determined in the unfiltered samples by flow injection analysis with chemiluminescence detection. Cobalt was measured at sea on unacidified samples using gallic acid as the luminescent reagent¹⁰. Copper was determined in the laboratory on acidified samples (pH 3) using a modification of the 1,10-phenanthroline method¹¹. The detection limits (two standard deviations) were 8 pM and 0.4 nM, respectively. Pore water samples were also obtained for cobalt analyses from sediments collected with a box corer. The pore water samples were extracted by centrifugation of sediments subsampled in a nitrogen-filled glove bag and filtered through $0.45 \mu\text{m}$ Nuclepore filters⁷. These samples were diluted with seawater containing a low concentration (10 pM) of cobalt before analysis by the same method used for the seawater samples.

The concentrations of cobalt and copper, and associated hydrographic properties measured in the water column, are listed in Table 1. Copper concentrations throughout the water column in Santa Monica Basin are similar to values observed by others in the eastern Pacific Ocean at the same depths^{13,14}, including the weak surface maximum (Fig. 1b). The cobalt profile from the surface to 600 m (Fig. 1c) is similar to the few measurements reported in open ocean waters by others^{1,2,12}. A sub-surface maximum of 110 pM is present at 35 m. The surface concentration of cobalt increased from 37 to 98 pM over the five-day span between the first and third hydrocast, although the concentration at the subsurface maximum was nearly constant. The increase in cobalt at the surface was accompanied by a corresponding increase in nitrate and decreases in oxygen and temperature. These changes are suggestive of strong upwelling during the windy period between the two surface hydrocasts. Cobalt concentrations decrease rapidly with depth below the maximum to values of 30 pM between 200 and 600 m. This decrease in cobalt concentration is indicative of removal by scavenging onto particle surfaces.

The concentration of cobalt increases from values of 30 to 50 pM near the sill depth to a maximum 124 pM in the subsill waters (Fig. 1c), in contrast to the behaviour observed for copper. This depth-associated increase in cobalt concentration does not occur in samples collected over oxidized sediments^{2,12}. Particulate cobalt is unlikely to account for the increase, because unacidified samples were analysed. There is also little evidence for large increases in particulate manganese concentrations near the bottom in Santa Barbara Basin¹⁵. The increase in cobalt concentrations below the sill depth must be supported by a large cobalt flux from the anoxic sediments. The flux can be determined from the near-bottom cobalt gradient (1.0 pM m^{-1}) calculated from the equation in Fig. 1c and the vertical eddy diffusion coefficient (K_z). Estimates of K_z for the bottom water in this basin have a large range. A value of $7 \text{ cm}^2 \text{ s}^{-1}$ has been calculated

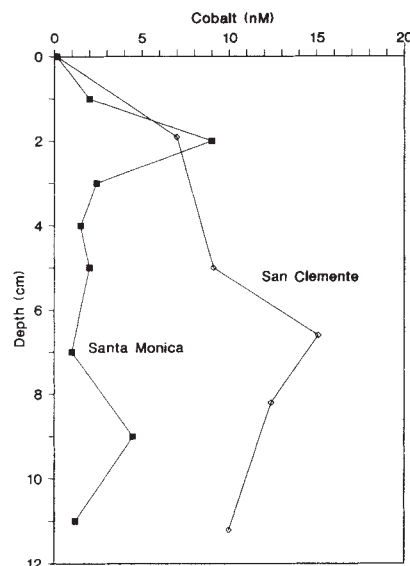


Fig. 2 Vertical profiles of cobalt in the upper 11 cm of the sediment measured in Santa Monica Basin. (■) This work; (◇) data from San Clemente Basin¹⁹.

from ^{222}Rn distributions¹⁶. We calculate a value of $1 \text{ cm}^2 \text{ s}^{-1}$ by extrapolating the K_z of $0.32 \text{ cm}^2 \text{ s}^{-1}$, measured near the sill with artificial tracers¹⁷, to the bottom. The extrapolation is accomplished by assuming an inverse dependence of K_z on the buoyancy gradient¹⁸. The estimates of K_z lead to estimated cobalt fluxes of 0.9 to 7 pmol cm^{-2} per day at the bottom.

Cobalt fluxes across the sediment-water interface can be calculated from its gradient in the porewaters of the sediments. Concentrations measured within the pore waters reach a maximum of 9 nM at a depth of 2 cm (Fig. 2). The profile in the upper sediments was similar to that observed in San Clemente Basin¹⁹, where there was a maximum concentration of 15 nM at 7 cm. The cobalt flux from the sediments, estimated from the 0 to 2 cm gradient (9,000 pM over 2 cm) and the molecular diffusion coefficient corrected for temperature, porosity and tortuosity ($0.3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ (ref. 19)), is 1.1 pmol cm^{-2} per day. The flux from the pore waters is similar to that estimated near the bottom of the water column.

Little of the cobalt flux escapes the subsill water of the basin. The estimated flux at sill depth is $0.08 \text{ pmol cm}^{-2}$ per day, calculated using the cobalt gradient and the eddy diffusion coefficient measured near the sill depth¹⁷. Cobalt diffusing upwards from the bottom must be rapidly scavenged and returned to the bottom. The scavenging rate can be estimated from the model

$$d(K_z \times d\text{Co}/dZ)/dZ - \alpha_z \text{Co} = 0$$

where α_z is the depth-dependent scavenging rate constant. We determined α_z by fitting an empirical equation to the cobalt data (Fig. 1c) and then differentiating the equation. The scavenging residence time ($0.69/\alpha_z$) varies from 2 yr at the sill to 0.2 yr at 880 m. In the open ocean, scavenging residence times for a suite of metals³ range from 22 yr (thorium) to 385 yr (copper), all substantially longer than that of cobalt in Santa Monica Basin. Considerably shorter scavenging residence times for manganese (0.005 yr) and chromium (0.44 yr) were obtained at the oxygen-sulphide interface in Saanich Inlet, an anoxic fjord²⁰.

The cobalt carrier phase must be easily reduced. This phase is probably a manganese oxide, given the similarities of the cobalt and manganese profiles noted in the oceanic water column^{1,2} and in sediments¹⁹. The carrier phase for copper is not labile in oxygen-deficient areas. As suggested by others^{3,4},

organic phases must be the primary carrier. The sensitivity of the cobalt flux to bottom-water oxygen concentrations suggests that its concentration in preserved benthic foraminifera may be a good tracer of oxygen in ancient oceans.

This work was supported by the Office of Naval Research (K.S.J.) and the National Science Foundation (W.M.B.). We thank Carol Chin and Dennis O'Neil for assistance with sampling and nutrient analyses, and Douglas Hammond for reviewing the manuscript.

Received 26 October 1987; accepted 18 February 1988.

- Bruland, K. W. in *Chemical Oceanography* Vol. 8 (eds Riley, J. P. & Chester, R.) 157-220 (Academic, London, 1983).
- Knauer, G. A., Martin, J. H. & Gordon, R. M. *Nature* **297**, 49-50 (1982).
- Balistreri, L., Brewer, P. G. & Murray, J. W. *Deep Sea Res.* **28**, 101-121 (1981).
- Fischer, K. et al. *Geochim. cosmochim. Acta* **50**, 1535-1543 (1986).
- Froelich, P. N. et al. *Geochim. cosmochim. Acta* **43**, 1075-1090 (1979).
- Emerson, S. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 78-87 (American Geophysical Union, Washington, DC, 1985).
- Berelson, W. M., Hammond, D. E. & Johnson, K. S. *Geochim. cosmochim. Acta* **51**, 1345-1363 (1987).
- Emery, K. O. *The Sea off Southern California* (Wiley, New York, 1960).
- Savard, C. E., Bottjer, D. J. & Gorsline, D. S. *Am. Ass. Petrol. geol. Bull.* **68**, 1179-1192 (1984).
- Sakamoto-Arnold, C. M. & Johnson, K. S. *Analyt. Chem.* **59**, 1789-1794 (1987).
- Yamada, M. & Suzuki, S. *Analyt. Lett.* **17**, 251-263 (1984).
- Martin, J. M. *Trans. Am. Geophys. Union* **66**, 1291 (1985).
- Boyle, E. A., Slater, F. & Edmond, J. M. *Earth planet. Sci. Lett.* **37**, 38-54 (1977).
- Bruland, K. W. *Earth planet. Sci. Lett.* **47**, 176-198 (1980).
- Shiller, A. M., Gieskes, J. M. & Price, N. B. *Geochim. cosmochim. Acta* **49**, 1239-1249 (1985).
- Berelson, W. M. thesis, Univ. Southern California, 1985.
- Ledwell, J. R., Watson, A. J. & Broecker, W. S. *Nature* **323**, 322-324 (1986).
- Gargett, A. E. *J. mar. Res.* **42**, 359-393 (1984).
- Heggett, D. & Lewis, T. *Nature* **311**, 453-455 (1984).
- Emerson, S., Cranston, R. E. & Liss, P. S. *Deep Sea Res.* **26**, 859-878 (1979).

Diatom and chemical evidence for reversibility of acidification of Scottish lochs

R. W. Battarbee, R. J. Flower, A. C. Stevenson, V. J. Jones*, R. Harriman† & P. G. Appleby‡

* Palaeoecology Research Unit, University College London, 26 Bedford Way, London WC1H 0AP, UK

† Freshwater Fisheries Laboratory, Pitlochry, Perthshire PH16 5LB, UK

‡ Department of Applied Mathematics and Theoretical Physics, University of Liverpool, Liverpool L69 3BX, UK

The cause-effect relationship between acid deposition and lake acidification is well established. Attention now focuses on the response of acidified lakes to reductions in SO₂ emissions¹. In the United Kingdom there has been a roughly 40% decline in national SO₂ emissions since the maximum in 1970 (ref. 2), and reductions in both the concentration of non-marine sulphate in precipitation^{2,3} and the wet deposition of sulphate² have been recorded in Scotland since the mid-1970s. The trend is especially clear at Eskdalemuir, an international monitoring site² (Fig. 1), and a site close to acidified lakes in Galloway, southwestern Scotland^{4,5}. Two of these lakes have been resampled. Diatom analysis of sediment cores taken between 1981 and 1986 show a trend towards progressively decreasing acidity in the uppermost sediments for both sites. Water chemistry records show that there have been significant decreases in both proton and sulphate concentrations since 1978.

We present data for Loch Enoch and the Round Loch of Glenhead. We have chemical records for both sites from repeated surveys at 2-3 month intervals in 1978-79 and 1984-86 and we have diatom data for two cores for Loch Enoch (sampled in 1982 and 1986) and three cores for the Round Loch of Glenhead (sampled in 1981, 1984 and 1986). Site details are published elsewhere⁶.

Table 1 Physical and chemical characteristics for Loch Enoch and Round Loch of Glenhead (1978-79 and 1984-86)

	Loch Enoch		Round Loch of Glenhead	
Altitude (m)	493		299	
Lake area (ha)	50		12.6	
Catchment area (ha)	186		95	
Catchment: lake	3.7		7.5	
Max. depth (m)	36		13.5	
	1978-79 (n = 6)	1984-86 (n = 6)	1978-79 (n = 5)	1984-86 (n = 6)
pH	4.40 (4.54)	4.62	4.65 (4.67)	4.77
H ⁺	40	24	22	17
Conductivity (μS cm ⁻¹)	38	40	36	41
Na ⁺	101	168	122	191
K ⁺	8	8	7	9
Ca ²⁺	15	27	27	41
Mg ²⁺	36	45	42	51
Cl ⁻	139	205	140	224
SO ₄ ²⁻	91 (89)	72	108 (112)	89
NO ₃ ⁻	10	15	6	9
Alkali	0	0	0	0
Absorbance (250 nm)	—	0.047	—	0.107
Total organic carbon (mg l ⁻¹)	— (1.1)	1.0	— (2.5)	2.2
Al (l) (μg l ⁻¹)	—	73	—	61
Al (nl) (μg l ⁻¹)	—	15	—	36
Al (total soluble μg l ⁻¹)	114 (150)		139 (165)	

Unless otherwise indicated values are expressed as microequivalents per liter (μeq l⁻¹). Catchment area excludes lake area. Aluminium values are shown for labile (l) and non-labile (nl) fractions. Values for 1979 from ref. 19 and indicated in parentheses.

Sampling and analytical methods used for water chemistry are described in ref. 7. The most abundant cations were determined by atomic absorption spectroscopy. Before 1986 anions were determined by an ion exchange method⁸, and afterwards by ion chromatography. Both methods were used for April 1986 samples and there was no significant difference between the results from the two methods. Total monomeric and labile forms of aluminium were measured by the pyrocatechol violet technique^{9,10}.

Sediments were sampled using Mackereth and Kajak corers and cores were sectioned at 0.5 cm intervals. ²¹⁰Pb dates were derived for all cores except for the 1984 Round Loch core, which was dated by biostratigraphic correlation with the 1981 core¹¹. Diatom analysis used standard procedures¹² and correspondence analysis used the program CANOCO¹³.

A comparison of the water chemistry of Loch Enoch between 1978-79 and 1984-86 (Table 1) indicates a rise in the mean pH of 0.22 pH units (H⁺ concentration 16 μeq l⁻¹, from 4.40 ± 0.12 (95% confidence limits) to 4.62 ± 0.04). The sample means for H⁺ were significantly different between the two periods (*P* = 0.05). Mean excess sulphate values were also significantly different (*P* = 0.05), falling from 77 ± 10 to 52 ± 3 μeq l⁻¹ (Table 1).

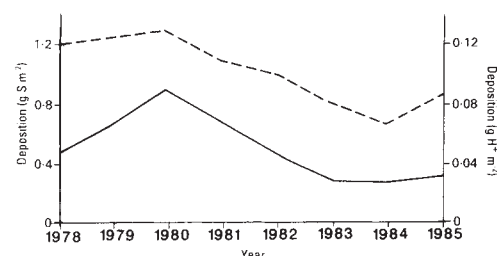


Fig. 1 Annual wet deposition of non-marine sulphate (broken line) and hydrogen ion from 1978-1985 for Eskdalemuir (from ref. 2).

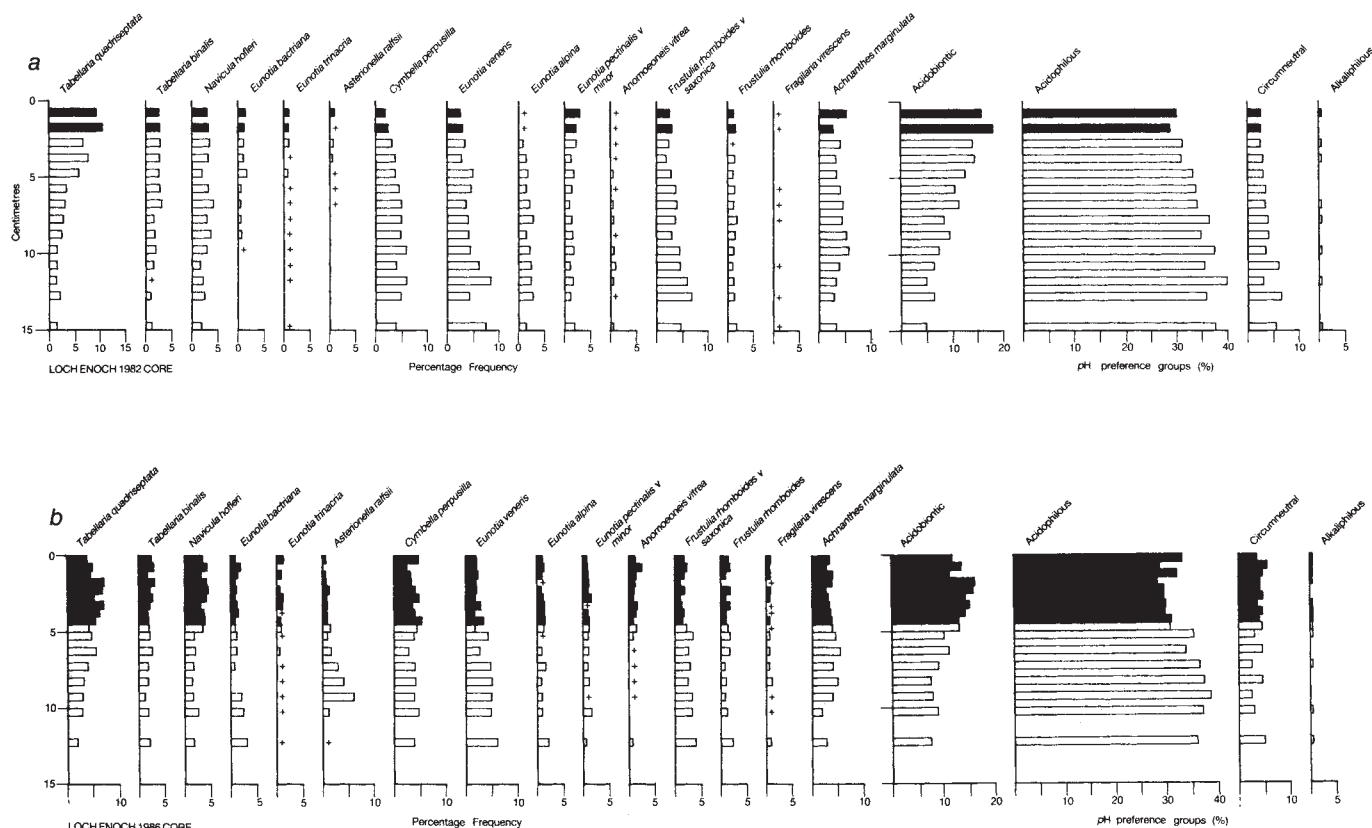


Fig. 2 Summary diatom diagrams for two cores from Loch Enoch: *a*, 1982 and *b*, 1986. Post-1970 levels are shaded.

The Round Loch of Glenhead (Table 1) showed no statistically significant difference ($P=0.05$) in mean pH and H^+ between the two dates, although the mean pH increased from 4.65 ± 0.16 to 4.77 ± 0.05 , a decline in proton concentration of $5 \mu\text{eq l}^{-1}$. Mean excess sulphate concentrations were significantly different, falling from $94 \pm 12 \mu\text{eq l}^{-1}$ to $67 \pm 14 \mu\text{eq l}^{-1}$.

Diatom diagrams for the upper 15 cm of the two cores from Loch Enoch are shown in Fig. 2. A general comparison between the assemblages of the two cores shows a good agreement, although the 1986 core has a peak of *Asterionella ralfsii* between 8 and 10 cm which is absent from the 1982 core. The 1986 core has a somewhat faster accumulation rate. ^{210}Pb dates show that the 1970 level for the 1982 and 1986 cores occurs at ~ 2.5 and 4.5 cm respectively, and ^{134}Cs measurements show the presence of post-Chernobyl radioactive fallout in the 1986 core. The 1982 core shows that the uppermost centimetre, representing sediment accumulation from about 1977–1982, contains lower proportions of acidobiontic species, whereas in the 1986 core the uppermost 1.5 cm (1981–86) has lower proportions of these species.

Although diatom changes in sediment cores are often used to reconstruct pH trends^{14,15}, the changes described here (0.1–0.2 pH units) are less than the errors in currently available models for pH reconstruction. The extent to which a change in trend has occurred can be evaluated more sensitively using ordination techniques that allow the degree of similarity between samples in a core to be shown. If the samples are linked in stratigraphic order, the series forms a time-track. In this case, correspondence analysis, an ordination method most suited to biological data, has been used. In Fig. 3 it can be seen that except for the deviation on axis 2 caused by the *A. ralfsii* peak, the two cores follow similar time-tracks until most recent times. In the 1986 core there is a clear floristic reversal and the surface

sample (~ 1982 –86) has an almost identical composition to samples at 4–5 cm (1968–1972).

The diatom data for the Round Loch of Glenhead have been treated similarly. Correspondence analysis time-tracks for the three cores are shown in Fig. 4. There is no evidence for reversal at the top of the 1981 core, but both the 1984 and 1986 cores show clear floristic reversals. The 1984 core has an accumulation

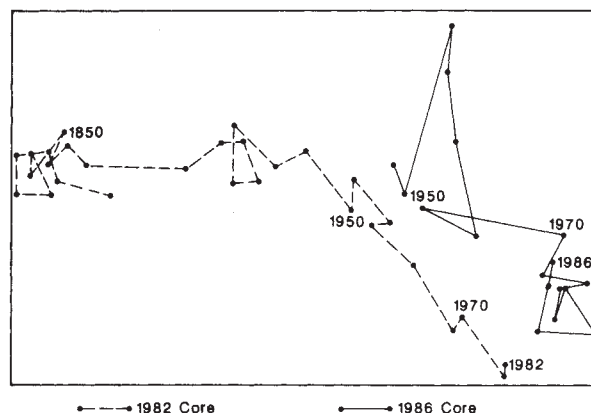


Fig. 3 Comparison of diatom assemblages from a 1982 (dashed line) and 1986 (continuous line) core from Loch Enoch using Correspondence Analysis¹³. Only axes 1 and 2 are shown. Successive stratigraphic levels are linked together to indicate the time-track of the individual core. Pre-1950 samples for the 1982 core show a high degree of similarity. Post-1950 samples show a directional change along axis 1. The 1986 core begins at about 1950 and shows a similar time-track, but the uppermost samples show a high degree of similarity and then track in the reverse direction.

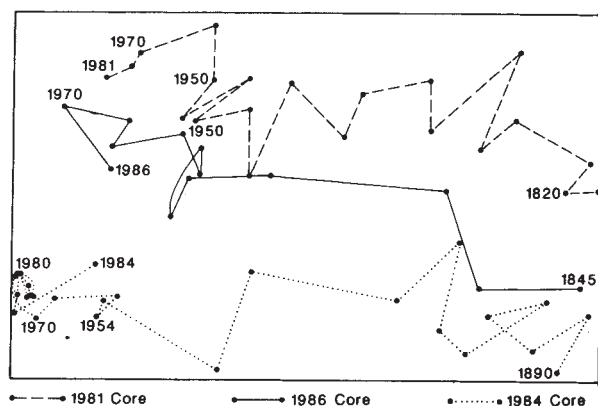


Fig. 4 Comparison of diatom assemblages from three cores from the Round Loch of Glenhead using Correspondence Analysis¹³. The 1981 core (dashed line) shows no indication of reversibility although post-1970 samples are quite similar. But the 1984 (dotted line) and 1986 (continuous line) cores show reversed time-tracks in the uppermost levels.

rate more than twice that of the 1986 core, allowing the recent history of the lake to be recorded with finer resolution. The data for this core show that little change occurred in the 1970s, indicating no further acidification, and that the improvement dates from about 1980. The uppermost samples of the 1984 and 1986 cores from Round Loch have a diatom composition similar to that of the 1950s.

Loch Enoch and Round Loch of Glenhead are still highly acidified lakes, but the data presented here show that no further acidification has taken place at either site since the mid-1970s, and that a small improvement has taken place since 1980. We cannot say whether concentrations of aluminium or total organic carbon have changed between the sampling periods as these were not measured in the 1978 survey.

Current diatom assemblages are not dissimilar from those recorded in the sediment between 1950 and 1970. As the period of rapid acidification at these sites occurred between 1930 and 1970 (ref. 6), it might be argued that a further significant reduction in acid deposition in Scotland could lead to a relatively larger response in water quality with the re-emergence of circum-neutral diatom taxa, such as *Anomoeoneis vitrea* and *Achnanthes microcephala*, in the diatom flora.

This trend towards improved conditions has also been observed at nearby Loch Valley (1978–9, pH 4.47, $n = 6$; 1984–6, pH 4.75, $n = 6$), (16) and cores taken in 1985 at several sites with moorland catchments in other parts of Scotland show diatom reversals in the uppermost sediment (unpublished observations; also ref. 17).

These data suggest that there is little delay (~10 yr) in the response of Galloway lakes to a decrease in acid deposition. Lakes have not continued to acidify as suggested by some models¹⁸, and some have improved. Indeed ordination of the core data suggests that the trajectory of the improvement is not only in the reverse direction but, in the case of diatoms, occurs without hysteresis.

We thank the Forestry Commission, John Thornicroft and Central Television, Stuart Phethean, Simon Patrick, Annette Kreiser and Sarah Wells for help with field work, and John Birks for discussion. This work was supported by the Royal Society (SWAP Project), The Natural Environmental Research Council, the Central Electricity Generating Board, and the Department of Environment (UK).

Received 16 September 1987; accepted 29 January 1988.

1. Dillon, P. J., Reid, R. A. & de Grosbois, E. *Nature* **329**, 45–48 (1987).
2. Barrett, C. F. *et al. Acid deposition in the United Kingdom, 1981–1985* (Warren Spring Laboratory, Stevenage, 1987).

3. Harriman, R. & Wells, D. E. *J. Wat. Pollut. Cont.* **84**, 215–222 (1985).
4. Flower, R. J. & Battarbee, R. W. *Nature* **305**, 130–133 (1983).
5. Battarbee, R. W., Flower, R. J., Stevenson, A. C. & Rippey, B. *Nature* **314**, 350–352 (1985).
6. Flower, R. J., Battarbee, R. W. & Appleby, P. G. *J. Ecol.* **75**, 797–824 (1987).
7. Harriman, R., Morrison, B. R. S., Caines, L. A., Collen, P. & Watt, A. W. *Wat. Air & Soil Pollut.* **32**, 89–112 (1987).
8. Mackereth, F. J. H., Heron, J. & Talling, J. F. *Freshwater Biol. Assn* **36** (1978).
9. Seip, H. M., Muller, L. & Naas, A. *Wat. Air & Soil Pollut.* **23**, 81–95 (1984).
10. Driscoll, C. J. Jr, Baker, J. P., Bisogni, J. J. & Schofield, C. *Nature* **284**, 161–164 (1980).
11. Jones, V. J. thesis, Univ. London (1987).
12. Battarbee, R. W. in *Handbook of Holocene Palaeoecology & Palaeohydrology* (ed. Berglund, B. E.) 527–570 (Wiley, Chichester, 1986).
13. Ter Braak, C. S. F. *Ecology* **67**, 1167–1179 (1986).
14. Battarbee, R. W. *Phil. Trans. R. Soc.* **305**, 451–477 (1984).
15. Flower, R. J. *Hydrobiologia* **143**, 93–103 (1987).
16. Harriman, R. & Wells, D. E. in *Acid Rain: Scientific and Technical Advances* (eds Perry, R., Harrison, R. M., Bell, J. N. B. & Lester, J. N.) 287–292 (Selper, London, 1987).
17. Flower, R. J., Darley, J., Rippey, B., Battarbee, R. W. & Appleby, P. G. *J. appl. Ecol.* (in the press).
18. Cosby, B. J., Wright, R. F., Hornberger, G. M. & Galloway, J. N. *Wat. Resour. Res.* **21**, 1591–1601 (1985).
19. Wright, R. F. & Henriksen, A. *Regional Survey of Lakes and Streams in Southwestern Scotland* Internal Report 72/70 SNSF Project, Oslo, Norway.

Evidence for a new sand transport process from experiments on Saharan dunes

J. Hardisty & R. J. S. Whitehouse

Marine Morphodynamics Unit, Department of Geography, Royal Holloway and Bedford New College, London University, Egham, Surrey TW20 0EX, UK

Sediment transport functions relate the rate of movement of granular solids per unit width of bed to some characteristics of the shearing fluid^{1–3}. Many have been applied to sand transport over aeolian dunes^{4–6}; these were developed for horizontal beds, however, whereas in natural environments there are often appreciable slopes on transport paths⁷. Theoretical analyses have usually been adopted^{8–10} to account for the effect of bedslope on the threshold of sediment transport^{11,12} and on the transport rate¹³. We report results from 462 experiments in the north-west Saharan sand seas in which we measured the sand transport rate for a range of wind speeds on a variety of up- and down-sloping dune surfaces. The results confirm the theoretical analyses for threshold criteria, but reveal transport rates related to the seventh power of the bedslope, considerably greater than those predicted by existing theory. This suggests that when grains land on the surface, they induce surface gravity flow in a process which has not previously been fully recognized.

The mechanics of the bedload process have been studied^{4,14,15} and the rate process modelled for horizontal beds^{4,5} using a Bagnold-type function for the bedload transport rate³.

$$j_b = k(u^2 - u_{cr}^2)u \quad (1)$$

where j_b is the mass transport rate per unit width, u is the wind speed at some characteristic height above the bed, u_{cr} is the threshold velocity for sediment movement and the coefficient of proportionality (k) is related to the properties of the fluid and the sediment. On sloping beds the function is corrected¹¹:

$$j_b = A k(u^2 - B^2 u_{cr}^2)u \quad (2)$$

where $A = k_b/k_0$ and $B = u_{crb}/u_{cro}$ are the ratios of the sloping-bed to flat-bed values (equal to unity on a horizontal surface). Theoretical analyses of the effect of bed slope on the threshold, B , have been presented by Allen^{16–18}, giving:

$$B = \frac{u_{crb}}{u_{cro}} = 1.373 \sin^{1/2}(i - b) \quad (3)$$

where i is the angle of internal friction of the sediment, here given the observed natural angle of repose of the desert dune

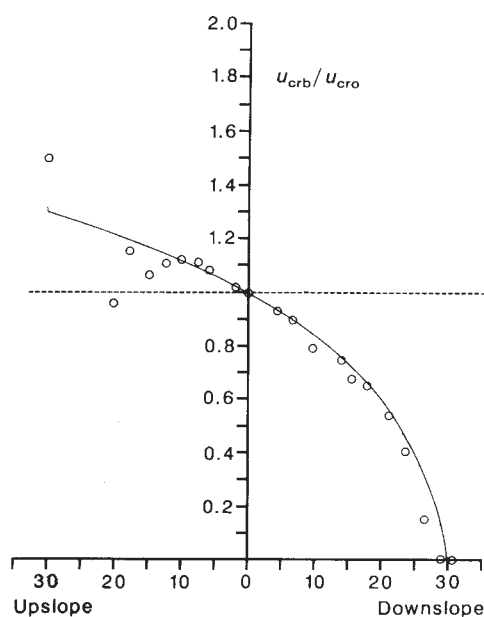


Fig. 1 Theoretical models of Allen¹⁶⁻¹⁸ and Dyer¹² (solid line) for the effect of bedslope on threshold, B (equations (3) and (4)). A single line represents both models because the differences are small; for example, for $b = 30^\circ$, $B = 1.2901$ (equation 3) versus $B = 1.2908$ (equation 4). The experimental data (O) are shown for comparison.

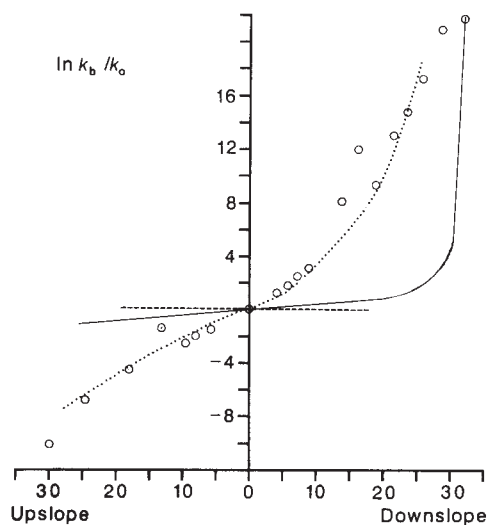


Fig. 2 Theoretical model of Bagnold¹³ (solid line) for the effect of bedslope on the transport rate, A (equation 5). The experimental data (O) and the empirical equation (6) are also shown (dotted line).

sand, 32° and b is the bedslope. An alternative analysis by Dyer¹² gives, in terms of threshold velocities:

$$B = \frac{u_{crb}}{u_{cro}} = \sqrt{\frac{\tan i - \tan b}{\tan i}} \cos b \quad (4)$$

These equations are very similar and are plotted in Fig. 1. The usual analysis for the effect of slope on the transport rate itself, A , follows that of Bagnold^{8,10,13}, giving:

$$A = \frac{k_b}{k_o} = \frac{\tan i}{\cos b (\tan i - \tan b)} \quad (5)$$

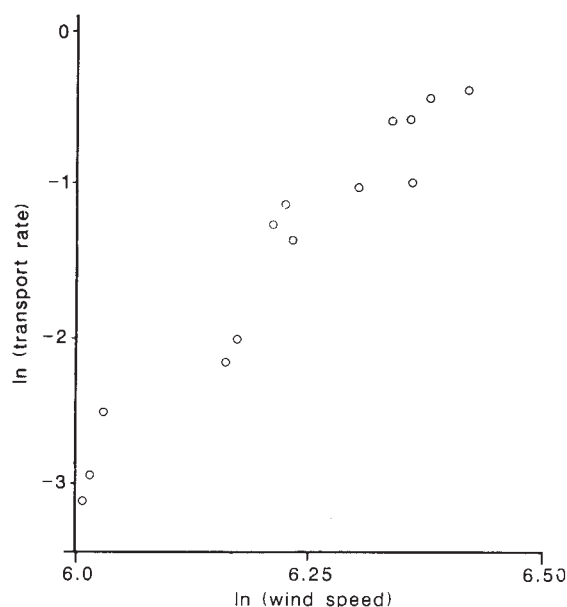


Fig. 3 A typical data set (bedslope of $b = +6^\circ$) with the windspeed in cm s^{-1} and the transport rate in $\text{g cm}^{-1} \text{s}^{-1}$.

Methods. A portable wind tunnel was used to measure the sand transport rate. The tunnel sides and top were fabricated from perspex and two battery powered fans delivered air into the parallel sided working section. The tunnel was placed on the dune surface and pushed vertically downwards so that the top was parallel with the bed. A trap, similar to that shown in Bagnold's⁴ Fig. 24, extended from the sand surface to the roof of the tunnel and was used to collect the total volume of sand moved per unit width. A hot-wire thermo-anemometer measured the wind velocity 5 cm above the bed in the tunnel. Separate measurements revealed a semi-logarithmic wind velocity profile between ~ 5 –25 mm above the bed; wind velocity was relatively constant above 25 mm. The wind velocity was increased over the sand surface to a constant velocity in excess of threshold, and the sand transport rates were determined for a wide range of velocities up to $\sim 6 \text{ m s}^{-1}$.

This equation is plotted in Fig. 2.

Equations (3), (4) and (5) were tested experimentally using a portable wind tunnel to measure the transport of a naturally well sorted fine sand in an area of desert dunes in the north-western Algerian Sahara. Battery-powered fans were used to generate wind over the sand surface and the transport rates were determined for a wide range of velocities in excess of threshold up to $\sim 6 \text{ m s}^{-1}$. Experiments were run for flat bed conditions and for a range of bedslopes, both upslope in the direction of flow (positive) and downslope (negative) to $\pm 32^\circ$. A typical data set is shown for $b = +6^\circ$ (Fig. 3). Analysis of the data for threshold velocities and the coefficient k on all slopes produced values of u_{crb} and k_b , from which A and B were determined. The results are compared with the theories presented above in Figs 1 and 2. Values determined for u_{cro} and k_o were 94.3 cm s^{-1} and $1.1 \times 10^{-13} \text{ g cm}^{-4} \text{ s}^2$ respectively. These are in accord with previously quoted values.

The results appear to support the theoretically derived values for the criterion, B (Fig. 1), although the data become scattered for slopes steeper than $b = +15^\circ$. The theory predicts that u_{crb} would tend to zero on negative slopes equal to i . The data, however, suggest that u_{crb} becomes zero on negative slopes at an angle 2 – 4° less than the natural angle of repose. Despite this, the use of equation (3) or (4) appears to provide a reasonable prediction for the slope-inclusive thresholds. The theoretical models are based upon fluid rather than impact thresholds, but the two coincide for the fine sand-size on the Saharan dunes (ref. 4, p194). The results for the transport rate k_b/k_o (Fig. 2) however, appear, to diverge considerably from theory (equation

5). Curve-fitting suggests that the transport rate is modified by the slope in accordance with:

$$A = \frac{k_b}{k_0} = \left[\frac{\tan i}{\tan i - \tan b} \right]^7 \quad (6)$$

This relationship is plotted as the dotted line in Fig. 2. We emphasize that equation (6) is an empirical estimate of A which suggests a much stronger dependence of the transport rate on the surface slope than that predicted by the theoretical equation (5). Our detailed observations of the sand-surface during experiments on the steeper bedslopes may explain this discrepancy. We observed that a wind blowing upslope transported most grains as surface-creep or saltation in the direction of the wind. Individual grains, however, vibrated rapidly about their stationary niche-position on the surface of the bed. Some of these grains rose a sufficient distance to clear the adjacent, downslope grains and rolled or slid down the slope in the opposite direction to the general movement. We have called this process impact-induced gravity flow because it appears that descending saltation grains transfer energy to the bed grains, some of which become energized, vibrate out of their niche-positions and move downslope under purely gravitational forces in the still-air region close to the bed. The enhanced downslope transport which results from this process may also explain the observation that u_{crb} becomes zero on negative slopes at angles less than i . Likewise, the increasingly complex dynamics which must operate when the impact induced gravity flow transports sand in the opposite direction to the wind may explain the scatter in the u_{crb} results for $b > +15^\circ$.

Sand gravity flow can be achieved, although with a different energy source, by placing loose sand on a sheet of sandpaper laid on a loudspeaker cone. Tilting the cone and the sheet slightly, with the loudspeaker switched off, does not cause the grains to move downslope because they are trapped in hollows on the sheet. The introduction of even a low-power signal into the loudspeaker produces downslope transport because the vibrating grains leave their niches and are free to move under gravity. Trials suggest that a signal of between 50–100 Hz will start the movement of 0.1 cm quartz sand. The impact induced gravity flow results in decreased upslope and increased downslope transport, and may account for the divergence between the theoretical k_b/k_0 ratio and the results shown in Fig. 2. The process would also agree with Anderson's¹⁹ observations of bed dilatation and the experiments on grain reptation by Mitha *et al.*¹⁵. The confirmation of the threshold criterion, B , concurs with our earlier subaqueous experiments²⁰ and with Howard's²¹ explanation for the difference between the orientation of small scale bedforms and wind direction.

We should like to thank the other members of Sahara '87 and C. P. Green for help in this work.

Crystal settling in a vigorously convecting magma chamber

Daniel Martin & Roger Nokes*

Research School of Earth Sciences, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia

* Present address: Department of Theoretical and Applied Mechanics, School of Engineering, University of Auckland, Private Bag, Auckland, New Zealand

There has been much debate concerning the mechanism of fractional crystallization in magma chambers. The traditional hypothesis of crystal settling has been widely replaced by the concept of *in situ* crystallization coupled with compositional convection. Observations from layered intrusions, however, are equivocal^{1–4}. Doubts have been raised about crystal settling on theoretical grounds because convective velocities in magma chambers are often much greater than the crystal settling velocities predicted by Stokes' law⁵, but there has been no experimental study of crystal settling in such vigorous convection. Here we present physical considerations and laboratory experiments which show that the phenomenon of particle settling in these conditions can be accounted for by a simple theory. Application of this theory to crystal settling in magma chambers suggests that crystal settling may be an efficient differentiation mechanism, at least in basaltic magma chambers, despite large convective velocities.

Particle settling in steady cellular convection has been studied previously^{6–8} but, given the high Rayleigh numbers (Ra) predicted for magma chambers^{9,10}, and the fact that even for infinite Prandtl-number fluids time-dependent behaviour occurs above $Ra \sim 10^9$ (ref. 11), it is considered that magma chambers usually undergo highly disordered 'turbulent' convection. Therefore, these works are not directly applicable to crystal settling in most magma chambers. There are also studies of settling in a turbulently convecting fluid^{12,13}, but these have used a theoretical approach that takes no account of the decrease in convective velocities as the boundaries are approached, a feature which, as discussed below, is of vital importance.

First we present some simple physical reasoning. Consider the case of particles of uniform size, suspended in a fluid undergoing unsteady convection driven by cooling from above, when the ratio, S , of the Stokes' law settling velocity, v_s , to the mid-depth r.m.s. (root mean square) vertical component of convective velocity, w_{rms} , is much less than one. Although the convective velocity is height-dependent^{14,15} and must decrease to zero at the boundaries, throughout most of the depth of the fluid the settling velocity is greatly exceeded by the convective



Fig. 1 Photograph of an experiment. The polystyrene particles show up as white dots. The clock at the left shows the experiment number (18) and the time in minutes and seconds. The large white blob in the centre of the photograph is a thermistor.

Received 14 December 1987; accepted 18 February 1988.

1. Bagnold, R. A. *Prof. Pap. U.S. geol. Surv.* 422-1 (1966).
2. Yalin, M. S. *Mechanics of Sediment Transport* 2nd edn (Pergamon, Oxford, 1977).
3. Hardisty, J. J. *Sedim. Petrology* 53, 1007-1010 (1983).
4. Bagnold, R. A. *The Physics of Blown Sand and Desert Dunes* (Methuen, London, 1941).
5. Sarre, R. D. *Prog. Phys. Geogr.* 11, 157-182 (1987).
6. Inman, D. L., Ewing, G. C. & Corliss, J. B. *Bull. geol. Soc. Am.* 77, 787-802 (1966).
7. Allen, J. R. L. in *Sedimentology: Recent Advances and Applied Aspects* (eds Brenchley, P. J. & Williams, B. P. J.) 7-28 (Blackwell, London, 1985).
8. Howard, A. D., Morton, J. B., Gad-El-Hak, M. & Pierce, D. B. *Sedimentology* 25, 307-338 (1978).
9. Richards, K. J. *J. Fluid Mech.* 99, 597-618 (1980).
10. Huthnance, J. M. *Estuar. Coast. Shelf Sci.* 15, 277-299 (1982).
11. Allen, J. R. L. *Mar. Geol.* 26, 281-328 (1980).
12. Dyer, K. R. *Coastal and Estuarine Sediment Dynamics* (Wiley, Chichester, 1986).
13. Bagnold, R. A. *Phil. Trans. R. Soc.* 249, 235-297 (1956).
14. Owen, P. R. J. *Fluid Mech.* 20, 225-242 (1967).
15. Mitha, S., Tran, M. Q., Werner, B. T. & Haff, P. K. *Acta mech.* 63, 267-278 (1986).
16. Allen, J. R. L. *Mar. Geol.* 48, 31-49 (1982).
17. Allen, J. R. L. *Mar. Geol.* 48, 51-73 (1982).
18. Allen, J. R. L. *Mar. Geol.* 48, 321-336 (1982).
19. Anderson, R. S. *Sedimentology* 34, 943-956 (1987).
20. Whitehouse, R. J. S. & Hardisty, J. *Mar. Geol.* 75, 135-139 (1988).
21. Howard, A. D. *Bull. geol. Soc. Am.* 88, 853-856 (1977).

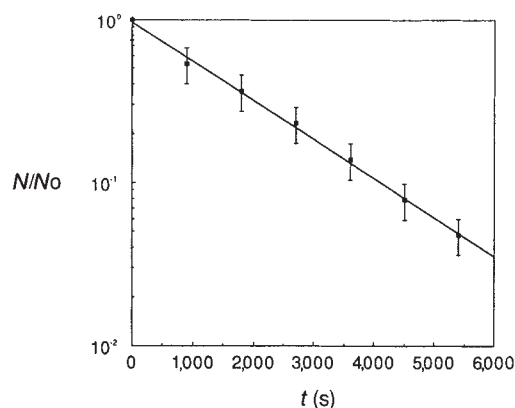


Fig. 2 Graph of ratio, N/N_0 , of number of particles in suspension at time t to number of particles in suspension at time $t=0$, against time, t , confirming exponential decay law (equations (4)).

velocity; thus, the particles behave very much like passive tracers and, because the convection is turbulent, they are very nearly uniformly distributed in the fluid. At the base of the fluid, however, where the convective velocity is zero, the particles must settle out with their full Stokes' law settling velocity. This implies that the decay in the number of particles in suspension in the convecting fluid occurs at a rate given by

$$\frac{dN}{dt} = -Av_s C(0) \quad (1)$$

where N is the number of particles in suspension, t is time, A is the area of the base of the fluid and $C(0)$ is the number density of particles at the bottom of the fluid.

As the particle distribution in the bulk of the flow is very nearly uniform, to a good approximation

$$C(0) = N/Ah \quad (2)$$

where h is the depth of the fluid, and equation (1) gives

$$\frac{dN}{dt} = -\frac{v_s}{h} N \quad (3)$$

Integrating equation (3) yields

$$N = N_0 \exp\left(-\frac{v_s t}{h}\right) \quad (4)$$

where N_0 is the initial number of particles in suspension. The number of particles in suspension decays with time according to an exponential law.

The analysis presented above is an approximate one and holds only if equation (2) is reasonable. Support for equation (2) comes both from the experiments described below, and from a more detailed theory to be presented fully elsewhere. The more general analysis considers the competing effects of particle settling and turbulent diffusivity in a way similar to Bartlett¹¹ and Huppert and Sparks¹², except that the variation of turbulent diffusivity (convective velocity) with depth is included. It can thus be shown that equation (2) holds to a very good approximation even when the settling velocity is comparable to the convective velocity over the whole depth of the fluid.

Further support for equation (4) and, implicitly, equation (2) comes from a series of laboratory experiments. Fluid in a 15 cm × 30 cm × 20 cm deep tank was cooled from above by circulating coolant through a heat exchanger which formed the roof of the tank. The tank was illuminated by slide projectors shining through two 1 cm-wide vertical slits on either side, so that

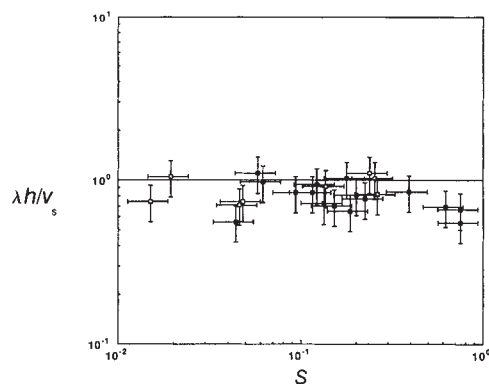


Fig. 3 Graph of dimensionless decay constant, $\lambda h/v_s$, against ratio, S , of settling velocity, v_s (calculated using Stokes' law for the middle of the particle size range), to mid-depth root mean square vertical component of convective velocity, w_{rms} . Solid boxes, experiments involving cooling from above only; open boxes, experiments involving cooling from above and heating from below.

unexpanded, approximately spherical polystyrene particles (measured density $1.033 \pm 0.001 \text{ g cm}^{-3}$) which had been introduced into the convecting fluid, were clearly visible. Three ranges of particle diameter were used: 0.21–0.31 mm, 0.31–0.42 mm and 0.42–0.50 mm. The fluids used consisted of water with various amounts of additives: NaCl was used to increase the density (up to 1.021 g cm^{-3}) and sodium carboxymethyl cellulose to increase the viscosity (up to 570 cSt ($5.7 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$)). The Rayleigh numbers appropriate to the experimental system ranged from 10^7 to 5×10^8 , well into the unsteady convection regime for these Prandtl numbers. Photographs were taken during the experiment (one such photograph is shown in Fig. 1) and the number of particles in suspension determined by counting the number of dots in a representative area of each photograph. Figure 2 shows the number of particles plotted against time for a typical experiment and confirms that an exponential decay law is obeyed. In addition to short exposure photographs, several 4s-exposure photographs were taken and the r.m.s. mid-depth vertical component of convective velocity was determined by measuring streak lengths. These streak photographs also revealed the structure of convection; in most experiments there were persistent quasi-cellular motions taking the form of one or two unsteady gyres. This is in keeping with studies of high Rayleigh number convection^{16–18} which demonstrate the existence of dominant scales of motion with horizontal length-scale similar to the depth of the fluid, or greater. In the experiments reported here the number of unsteady cells was restricted by the dimensions of the tank; at higher aspect ratio, more cells can be expected.

For each experiment a characteristic exponential decay constant, λ , (calculated from the slope of the line in Fig. 2) was determined. Figure 3 (solid boxes) presents the results of all of the experiments in which the dimensionless decay constant, $\lambda h/v_s$, is plotted against the ratio of velocities, $S = v_s/w_{rms}$, where w_{rms} was determined from streak photographs, and v_s is the settling velocity appropriate to a particle with a diameter in the middle of the range used. For $S < 0.5$, the decay constant is close to that predicted by the reasoning given above (namely $\lambda h/v_s = 1$). In fact the points appear to lie systematically below the line $\lambda h/v_s = 1$. The presence of a range of particle sizes is one reason for this. If the exponential decay law is applied to a range of particle size, then the total number of particles remaining in suspension at time t is given by

$$N(t) = \int_{x_1}^{x_2} \frac{N_0}{x_2 - x_1} \exp\left(-\frac{v_s t}{h}\right) dx \quad (5)$$

where a uniform distribution of particle sizes between radii x_1

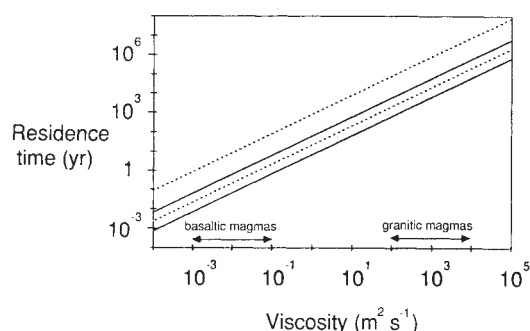


Fig. 4 Graph of residence time of spherical crystals plotted against viscosity for a 1-km-deep chamber. Solid lines, predictions for olivine crystals ($x = 0.05\text{--}0.15$ cm, $\rho_c = 3.3$ g cm $^{-3}$, $\Delta\rho = 0.6$ g cm $^{-3}$), where ρ_c is the density and x is the radius of the crystal. Dashed lines, predictions for plagioclase crystals ($x = 0.005\text{--}0.3$ cm, $\rho_c = 2.7$ g cm $^{-3}$, $\Delta\rho = 0.05$ g cm $^{-3}$). Also shown are the range of viscosities for basaltic and granitic magmas.

and x_2 has been assumed. Equation (5), on integration, gives

$$N(t) = Bt^{-1/2}[\text{erf}(kx_2t^{1/2}) - \text{erf}(kx_1t^{1/2})] \quad (6)$$

with

$$B = \frac{N_0}{x_2 - x_1} \sqrt{\frac{9\pi h\nu}{8g\Delta\rho}} \quad \text{and} \quad k = \sqrt{\frac{2g\Delta\rho}{9h\nu}}$$

where $\Delta\rho$ is the excess density of the particles, g is the acceleration due to gravity and ν is the kinematic viscosity of the fluid. For the duration of our experiments the deviation of this decay law from a straight line, when the logarithm of the number of particles is plotted against time, is negligible compared with experimental error, but the values of $\lambda h/\nu_s$ determined from the experiments should underestimate the correct values by $\sim 10\%$. For $S > 0.5$, the experimentally observed decay-constant may be significantly less than predicted; this is because the assumption of one-dimensionality, implicit in the theory, breaks down. During these experiments it was observed that particles were concentrated into the upwelling parts of the mean flow, so that the number-density of particles immediately above the bottom boundary was less than the average number-density of particles in the tank.

Experiments and theory therefore indicate that the rate of settling of particles in high-Ra convection driven by cooling from above is given by equation (3), and is independent of the convective velocity, provided the Stokes' law settling velocity of the particles is exceeded by a sufficient amount (namely $S < 0.5$). Below we apply equation (3) to magma chambers, but first we point out two potentially important limitations on the extension of our theory to crystal settling in magmas.

First it has been assumed that particle concentrations are low. High particle concentrations have two effects. Stokes' law is modified, and ν_s in equation (3) is changed¹⁹. Also, convection

may be modified by the presence of the particles. Little is known about this latter possibility, but it seems likely that the presence of high concentrations of particles will tend to dampen the turbulence of convective motions. The particle concentrations above which this effect becomes important are not known, neither are the crystal concentrations in magma chambers. The original particle concentration in the experiments reported here was usually around 0.3 vol.%.

Second it has been assumed that particles are not re-entrained into the convecting fluid once they have settled out. This was true in the experiments reported above. Subsequent experiments investigated the effect of driving convection by heating from below, as well as by cooling from above. In all experiments except one there was no re-entrainment. (The results of these experiments are plotted as open boxes in Fig. 3.) When the most viscous fluid (570 cSt (5.7×10^{-4} m 2 s $^{-1}$)) was used with the smallest particles, re-entrainment was observed. In this experiment the number of particles in suspension decayed with time until a steady state was reached when the rate of re-entrainment balanced the rate at which particles settled out. We do not know whether, and under what circumstances re-entrainment will occur in magma chambers. Certainly the possibility of re-entrainment is enhanced if compositional convection is being driven by further crystal growth on the floor, but further work is required to develop a re-entrainment criterion.

If equation (3) is accepted, it can be demonstrated that crystal settling could be an efficient differentiation mechanism in magma chambers. Since an exponential decay law is obeyed, we can calculate a half-life, or residence time, of typical crystals in typical magmas. This is given by $\ln 2 \times h/\nu_s$. Figure 4 shows the range of residence times expected for olivine and plagioclase crystals in 1-km-deep chambers as a function of magma viscosity. For olivine crystals in basaltic chambers typical residence times are a few days to a few years, whilst plagioclase crystals have residence times up to 100 years. These times are very short compared with the many thousands of years that a magma chamber takes to solidify, and on this timescale crystal settling is virtually instantaneous. In granitic chambers the range of residence times is from 100 years to 10 million years: crystal settling should be able to operate in the least viscous granitic magmas, but not in the most viscous ones. Ideally these calculations should take into account the growth of crystals whilst they are in suspension, but because this is so little understood, we cannot easily modify our calculated residence times. These 'first order' calculations are sufficient to demonstrate our main point. We have shown both theoretically, and by using a direct laboratory experiment, that, despite the fact that convective velocities in magma chambers are commonly orders of magnitude higher than the settling velocities of crystals, crystal settling (at least in basaltic chambers) is potentially a very efficient mechanism for achieving differentiation. If crystals occur in suspension in a convecting basaltic magma chamber, they will readily and rapidly settle out.

We thank Ian Campbell, Ross Griffiths and Stewart Turner for valuable discussions, and Derek Corrigan, Ross Wyld-Browne and Tony Beasley for technical assistance.

Received 23 September 1987; accepted 18 February 1988.

- Irvine, T. N. *Geol. Soc. Am. Mem.* **138** (1974).
- Campbell, I. H. *Lithos* **11**, 311–323 (1978).
- McBirney, A. R. & Noyes, R. M. *J. Petrol.* **20**, 487–554 (1979).
- Parsons, I. & Butterfield, A. W. *J. geol. Soc. Lond.* **138**, 289–305 (1981).
- Sparks, R. S. J., Huppert, H. E. & Turner, J. S. *Phil. Trans. R. Soc. A* **310**, 511–534 (1984).
- Stommel, H. *J. mar. Res.* **8**, 24–29 (1949).
- Marsh, B. D. & Maxey, M. R. *J. Volcan. geotherm. Res.* **24**, 95–150 (1985).
- Weinstein, S. A., Yuen, D. A. & Olson, P. L. *Earth planet Sci. Lett.* **87**, 237–248 (1988).

- Martin, D., Griffiths, R. W. & Campbell, I. H. *Contr. Miner. Petrol.* **96**, 465–475 (1987).
- Carrigan, C. R. *Geophys. Res. Lett.* **14**, 915–918 (1987).
- Jarvis, G. T. *Phys. Earth. planet. Inter.* **36**, 305–327 (1984).
- Bartlett, R. W. *Am. J. Sci.* **267**, 1067–1082 (1969).
- Huppert, H. E. & Sparks, R. S. J. *Contr. Miner. Petrol.* **75**, 279–289 (1980).
- Kraichnan, R. H. *Phys. Fluids* **5**, 1374–1389 (1962).
- Garon, A. M. & Goldstein, R. J. *Phys. Fluids* **16**, 1818–1825 (1973).
- Deardorff, J. W. & Willis, G. E. *J. Fluid Mech.* **28**, 675–704 (1967).
- Fitzgerald, D. E. *J. Fluid Mech.* **73**, 693–719 (1976).
- Krishnamurti, R. & Howard, L. N. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1981–1985 (1981).
- Barnea, E. & Mizrahi, J. *J. chem. Engng* **5**, 171–189 (1973).

Algal biomass unaltered by food-web changes in Lake Michigan

John T. Lehman

Department of Biology and Great Lakes Research Division, Natural Science Building, The University of Michigan, Ann Arbor, Michigan 48109, USA

Bythotrephes cederstroemii Schoedler (Crustacea: Cladocera), a predator previously confined to the Palearctic, has successfully invaded the North American Great Lakes¹⁻³. *Bythotrephes* is a voracious predator on herbivorous Cladocera, including the dominant grazers in Lake Michigan during the summer. Lake Michigan has been the source of active debate regarding the relative importance of nutrient income versus food-web relations to its trophic state and water quality⁴⁻⁶. The recent species invasion has directly altered the lake's food web at a middle trophic level. During summer 1987 *Bythotrephes* populations increased rapidly in the offshore regions of Lake Michigan and abundances of herbivorous zooplankton simultaneously declined. Despite the resulting relaxation of herbivory, particulate chlorophyll concentrations, an index of algal biomass, did not increase. These results suggest that primary producers are most constrained by abiotic forces in this deep oligotrophic lake.

The arrival of *Bythotrephes* in North America¹⁻³ has caused a novel perturbation of existing plankton communities. The rising abundances of *Bythotrephes* at an offshore reference station at 43° N 86°40' W (36-km offshore; depth, 100 m) in Lake Michigan were associated with pronounced changes in the abundances of other zooplankton taxa. Zooplankton abundances on 30 June 1987, when *Bythotrephes* was rare, in most cases equalled or exceeded abundances recorded during 1985 and 1986 before the arrival of the predator (Table 1). Epilimnion temperatures were warmer in 1987 than in the previous two years, and the warm water promoted the early success of most crustacean taxa other than *Limnocalanus* and *Senecella*, which are hypolimnion-dwelling cold stenothermic copepods⁷. By 22 July 1987 *Bythotrephes* had increased to more than 200 individuals m⁻², and the abundances of five other plankton species were significantly different from levels recorded in 1985 and 1986 (Table 1).

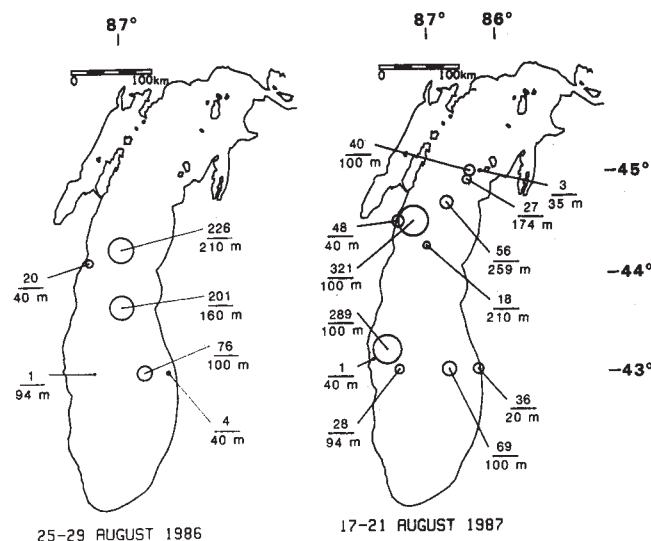


Fig. 1 Abundances (individuals m⁻²) of *Bythotrephes* in Lake Michigan during survey cruises in August 1986 and 1987. Areas of the plotted circles are proportional to abundances. Numerators, individuals m⁻² in water column from surface to within 5 m of bottom; denominators, depth of water column.

The depth-integrated plankton abundances reported in Table 1 have only a 20% estimation error⁸. By first-order error propagation⁹, abundance ratios between years are subject to standard errors of estimation equal to 28% of the ratio. Ratios are thus not regarded to be significantly different from 1.0 at the 0.05 probability level (1.0 ± 2 s.e.m.) unless they are less than 0.44 or greater than 2.27 (the reciprocal of 0.44). All three *Daphnia* species (*D. pulicaria*, *D. galeata mendotae* and *D. retrocurva*), as well as *Leptodora kindtii*, declined in July 1987 and subsequently remained low during sampling cruises in August and September. The only crustacean to increase substantially, other than *Bythotrephes* itself, was *Epischura lacustris*. *Epischura* is univoltine and its population growth was due to accelerated recruitment of copepodids from naupliar stages earlier than usual because of warm temperatures. *Epischura* is carnivorous only in its late copepodid stages, and although capable of consuming *Daphnia* occasionally, it preys principally on zooplankton much smaller than *Daphnia*¹⁰.

The spatial distribution of *Bythotrephes* during late August 1986 was consistent with a presumed invasion route from Lake Huron in the north (Fig. 1). By 1987, *Bythotrephes* was no longer in an early colonizing phase, but populations nonetheless developed asynchronously in different regions of the lake. In August 1987, *Bythotrephes* was in greatest abundance at deep (100 m) stations in western Lake Michigan. Populations on the east side of the lake had declined from similar high abundances one month earlier (Table 2).

Generation times for parthenogenically reproducing *Bythotrephes* are less than two weeks¹¹. Females are iteroparous, and clutch sizes averaged 4.4 embryos per female in 1987. With

Table 1 Abundances of dominant crustacean zooplankton at an offshore reference station in Lake Michigan (43° N 86°40' W) during comparable cruise periods in 1985-87

Taxon	Individuals m ⁻²			Ratios	
	1985	1986	1987	1987	1987
	(25-26 June)	(7-8 July)	(30 June)	1985	1986
<i>Diaptomus</i> spp. C1-C6	472,600	663,700	832,200	1.76	1.25
<i>Cyclops</i> spp. C1-C6	290,100	423,700	719,000	2.48	1.70
<i>Daphnia galeata mendotae</i>	17,800	84,280	97,460	5.48	1.16
<i>Daphnia pulicaria</i>	345	3,473	2,300	6.67	0.66
<i>Daphnia retrocurva</i>	32,400	56,480	46,170	1.42	0.82
<i>Bosmina longirostris</i>	30,800	70,360	18,540	0.60	0.26
<i>Limnocalanus macrurus</i> C1-C6	14,600	7,480	7,210	0.49	0.96
<i>Senecella calanoides</i> C1-C6	456	380	318	0.70	0.84
<i>Epischura lacustris</i> C1-C6	110	64	738	6.71	11.53*
<i>Leptodora kindtii</i>	169	1,975	1,388	8.21	0.70
<i>Bythotrephes cederstroemii</i>	0	0	11		
Temperature at 5 m (°C)	11.5	14.9	18.0		
	16-18 July	22-23 July	22 July		
<i>Diaptomus</i> spp. C1-C6	383,800	705,100	646,400	1.68	0.92
<i>Cyclops</i> spp. C1-C6	316,300	408,400	404,000	1.28	0.99
<i>Daphnia galeata mendotae</i>	28,040	67,360	2,260	0.08	0.03*
<i>Daphnia pulicaria</i>	18,800	17,440	345	0.02	0.02*
<i>Daphnia retrocurva</i>	15,740	18,380	16	0.00	0.00*
<i>Bosmina longirostris</i>	23,090	39,300	26,610	1.15	0.68
<i>Limnocalanus macrurus</i> C1-C6	9,550	9,260	15,520	1.62	1.68
<i>Senecella calanoides</i> C1-C6	456	292	512	1.12	1.75
<i>Epischura lacustris</i> C1-C6	522	340	3,842	7.32	11.30*
<i>Leptodora kindtii</i>	2,677	2,892	0	0.00	0.00*
<i>Bythotrephes cederstroemii</i>	0	0	239		
Temperature at 5 m (°C)	18.6	20.6	22.4		

Only copepodid stages (C1-C6) are tabulated for copepods. Abundances significantly different in 1987 from both previous years are denoted by an asterisk.

Table 2 Reproductive condition of adult female *Bythotrephes* collected from offshore Lake Michigan in 1987

	June	July	August	September
Fecund parthenogenic females	60	368	15	71
Barren females	10	200	8	31
Gametogenic females with resting eggs	0	72	12	31
<i>Bythotrephes</i> m ⁻²	11	239	69	164
Abundance ratios				
<i>Daphnia</i> : <i>Bythotrephes</i>	13,270	11	318	380

Frequency distributions change significantly through time (overall $\chi^2 = 52.8$, 6 d.f.; $P = 1.3 \times 10^{-9}$). There were proportionally more barren and gametogenic females in July than in June, and proportionally more gametogenic females in August than in July. Reproductive conditions of females in August and September were statistically indistinguishable.

its short generation time, *Bythotrephes* can quickly dominate the dynamics of prey populations, but its influence may subside equally rapidly. The collapse of the *Daphnia* population in Lago Maggiore, for instance, has been attributed to predation from *Leptodora* and *Bythotrephes*¹²⁻¹⁴. Short-lived predators like *Bythotrephes*, however, may become victims of their own success. By mid-July 1987, the ratio of *Daphnia* to *Bythotrephes* had dropped to less than 30 to 1 at the offshore reference station (Table 2). The subsequent decline in *Bythotrephes* in August was associated with production of resting eggs, rather than parthenogenic eggs, an indication of food limitation. The proportions of barren *Bythotrephes* females and those bearing resting eggs increased significantly from June to July, when the *Daphnia* populations declined (Table 2). *Daphnia* species, in contrast, remained parthenogenic and fecund during their decline. Cladocera switch from parthenogenic reproduction to gametogenic production of resting eggs when food levels fall precipitously¹⁵.

Leptodora, another predatory cladoceran, was virtually absent by July 1987 during the peak population development of *Bythotrephes*. *Bythotrephes* co-exists with *Leptodora* in European lakes¹⁶, and so the loss of *Leptodora* in Lake Michigan was an unusual development. The two predators forage in different ways, but they both take small Cladocera as prey. *Bythotrephes* is the stronger swimmer and it possesses a stout (~1 cm) abdominal spine with up to three recurved barbs¹⁷. *Leptodora* is an almost transparent, soft-bodied animal with no sclerotized structures besides its mandibles. In June, *Bythotrephes* were found with remains of *Leptodora* trailing from their spines. Thus some incidental direct mortality may have caused the decline of *Leptodora*.

Nutrient income to Lake Michigan and per capita loading of phosphate by human activities have been restricted for more than a decade, in an effort to maintain the oligotrophic condition of the lake¹⁸. Owing to the century-long hydraulic residence time of the basin the responses have been relatively gradual, but evidence is accumulating that the measures are having success⁴. Lake Michigan has also been stocked with piscivorous salmon for two decades. Resulting changes in abundances of forage fish, primarily the planktivorous alewife (*Alosa pseudoharengus*) have been expected to cause changes that rival the effects of nutrient reduction. In Lake Michigan this dichotomy of management themes has initiated debate about the strength of nutrient controls and fishery management on trophic state and water quality⁴⁻⁶. Now, however, the lake ecosystem has endured a rapid manipulation at a middle trophic level by direct species invasion. Coincidentally, the invasion has provided evidence to refute the hypothesis that trophic interactions control the total biomass of algae in Lake Michigan. *Daphnia* are biomass dominants and the major epilimnetic herbivores during summer in Lake Michigan⁸, and alterations in their abundance have been linked to changes in algal concentrations in other lakes¹⁹. Figure 2 shows the mean concentrations of chlorophyll *a* from surface to 20 m at the reference

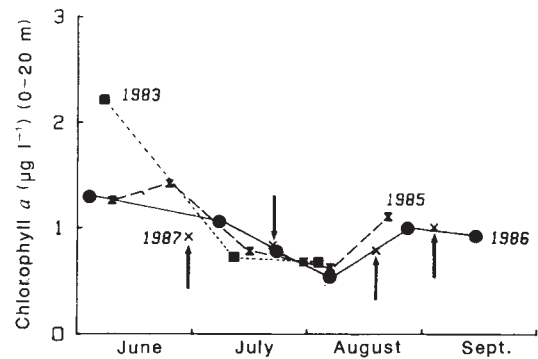


Fig. 2 Mean concentrations of chlorophyll *a* from surface to 20 m at the offshore reference station at Lake Michigan. Values in 1987 are indicated by arrows.

station in 1985, 1986 and 1987, as well as chlorophyll *a* measured at nearby stations in 1983 (ref. 14). During July and August 1987, when *Daphnia* populations were at 10–50-fold lower abundance than in previous years, the biomass of algae measured as chlorophyll *a* was unchanged from previous values. Evidently, the maximum biomass of algae in Lake Michigan is constrained by forces other than herbivory.

I thank J. Bowers, D. Burdette and the crew of the RV Laurentian for assistance with field sampling. This work was supported by the NSF.

Received 14 December 1987; accepted 25 February 1988.

1. Lehman, J. T. *Oecologia* **74**, 478–480 (1987).
2. Lange, C. & Cap, R. *J. Great Lakes Res.* **12**, 142–143 (1986).
3. Bur, M. T., Klarer, D. M. & Krieger, K. A. *J. Great Lakes Res.* **12**, 144–146 (1986).
4. Scavia, D., Fahnenstiel, G. L., Evans, M. S., Jude, D. J. & Lehman, J. T. *Can. J. Fish. Aquat. Sci.* **43**, 435–443 (1986).
5. Edgington, D. in *The Future of Great Lakes Resources* 24–33 (University of Wisconsin Sea Grant Institute, Madison, 1984).
6. Schelske, C. L., Stoermer, E. F., Conley, D. J., Robbins, J. A. & Glover, R. M. *Science* **222**, 320–322 (1983).
7. Balcer, M. D., Korda, N. L. & Dodson, S. I. *Zooplankton of the Great Lakes* (University of Wisconsin Press, Madison, 1984).
8. Dorazio, R. M., Bowers, J. A. & Lehman, J. T. *J. Plankton Res.* **9**, 891–899 (1987).
9. Meyer, S. L. *Data Analysis for Scientists and Engineers* (Wiley, New York, 1975).
10. Kerfoot, W. C. *Limnol. Oceanogr.* **22**, 316–325 (1977).
11. Mordukhai-Boltovskaya, E. D. *Doklady Akad. Nauk SSSR* **112**, 1133–1135 (1957).
12. Argentesi, F., de Bernardi, R. & Di Cola, G. *Mem. Ist. Ital. Idrobiol.* **31**, 245–275 (1974).
13. de Bernardi, R. *Mem. Ist. Ital. Idrobiol.* **31**, 221–243 (1974).
14. de Bernardi, R. & Canali, S. *Mem. Ist. Ital. Idrobiol.* **32**, 365–392 (1975).
15. d'Abramo, L. *Limnol. Oceanogr.* **25**, 422–429 (1980).
16. de Bernardi, R., Giussani, G. & Manca, M. *Hydrobiologia* **145**, 225–243 (1987).
17. Ischreyt, G. *Int. Rev. Ges. Hydrobiol.* **31**, 181–202 (1934).
18. Slater, R. W. & Bangay, G. E. in *Phosphorus Management Strategies for Lakes* (eds Loehr, R. C., Martin, C. S. & Rast, W.) 13–26 (Ann Arbor Science, 1980).
19. Edmondson, W. T. & Litt, A. H. *Limnol. Oceanogr.* **27**, 272–293 (1982).

Rapid changes in genetic structure of epidemic populations of *Ophiostoma ulmi*

C. M. Brasier

Forest Research Station, Alice Holt Lodge, Farnham, Surrey GU10 4LH, UK

The Northern Hemisphere is currently experiencing simultaneous epidemics of Dutch elm disease, caused by two new and highly pathogenic forms of *Ophiostoma* (*Ceratocystis*) *ulmi*, termed the Eurasian (EAN) and North American (NAN) races of the aggressive subgroup. The NAN aggressive has spread across North America^{1,2} and into western Europe, probably through Britain^{3,4}; the EAN has migrated westwards across Europe, and into south-western and central Asia^{4,5}. They are now replacing the 'old' non-aggressive subgroup believed to be responsible for the first milder epidemic of the disease during the 1920s to 1940s (refs 5–9), and most mature elms are likely to be killed in affected areas⁵. This situation has provided an opportunity to examine

Table 1 Structure of epidemic and post-epidemic population samples

Site and status	Sample size	Uniform or supergroup component			Second vc group		Heterogeneous component			
		% of sample	% A type*	% d-infected*	% of sample	% A type*	% of sample	vc group frequency†		% A type*
								No. of groups/no. of isolates	Ratio isolates/groups	
<i>a</i> , EAN aggressive sites										
Poland (1980) recent epidemic	52	49	0	31	8	0	43	19/20	1.05	39
Romania (1980) post-epidemic	71	20	0	7	7	0	73	22/26	1.18	12
<i>b</i> , NAN aggressive post-epidemic sites, southern Britain										
Chichester (1982)‡	40	15	0		—	—	85	12/13	1.08	15
Southampton area	72	19	0	0			82			22
Orsett (1983)‡	86	13	0	10	—	—	87	14/14	1.00	16
South Essex (1983)	71	10	0	0			90			9
Tewkesbury (1983)‡	67	16	0	18	16	0	67	9/10	1.11	13
Gloucestershire (1983)	83	7	0	0	10	0	83			7
Total or mean	419	13	0	6			82			14
<i>c</i> , NAN aggressive epidemic front sites, Spain										
Avila area (1985)	74	88	40	3	—	—	12	5/9	(1.80)	(44)
Guadalajara area (1985)	39	95	0	49	—	—	5	2/2	(1.00)	(50)

Each sample was from the twig of a different diseased elm, isolated as described in ref. 27. Vegetative compatibility and mating type¹⁰ tests were carried out on elm sapwood agar²⁷ at 20 °C in darkness. Vegetative incompatibility was determined on the basis of a clear wide or narrow reaction¹⁰. d-Infection frequencies were estimated for vc supergroup isolates only on the basis of ability to give a strong d-reaction^{16,17} in repeated tests against a healthy isolate of the same vc group, and are minimal estimates as they reflect only overt disease and do not account for isolation failures due to overt d-infection or reflect latent¹⁸ d-infection. Most d-infected isolates exhibited degenerate colony phenotypes^{16,17} whereas healthy isolates were normal.

* Assessed for all isolates.

† All × all pairings of a 20–90% random sample of the heterogeneous component. At the Polish, Avila and Guadalajara sites the non-aggressive subgroup still represented ~3, 32 and 69% of the overall *O. ulmi* sample respectively, whereas at the Romanian and British sites it was absent.

‡ Chichester, Orsett and Tewkesbury are the three original epidemic outbreak areas studied in 1971 (ref. 6).

changes in the genetic structure of a plant pathogen population during a catastrophic pandemic event. I present here evidence that current epidemic front populations of *O. ulmi* are largely clonal, and that a rapid change to heterogeneity then occurs which could be associated with the spread of mycoviruses in the pathogen clones.

The vegetative incompatibility system in *O. ulmi* is a multi-locus bi- or multiallelic 'self-non-self' recognition system, analogous to animal tissue incompatibility systems, which promotes the breakdown of hyphal fusions between adjacent mycelia. Whether two isolates are vegetatively compatible (vc) or incompatible, and hence of the same or of different vc groups, can be determined from their interaction in culture¹⁰. As *O. ulmi* is normally outcrossing with two mating types, A or B (ref. 10), the system can generate large numbers of vc groups, and the number of groups can be used as an indicator of population heterogeneity.

A preliminary analysis of geographically scattered isolates suggested that EAN and NAN aggressive populations each contained a single dominant vc group, the remainder being heterogeneous. The two dominant vc groups, a different group in each case, were termed the EAN and NAN vc 'supergroups'¹⁰. This distinctive population structure was confirmed in more detailed sampling at a recent EAN epidemic site in Poland, and at a post-epidemic site in Romania. The EAN supergroup comprised ~50 and ~20% of the samples, respectively, with a second major vc group occurring at lower frequency (~7%) at both sites (Table 1a). The remainder was highly heterogeneous for vc groups. Likewise at several NAN post-epidemic sites in Britain the NAN supergroup comprised ~13% of the population (Table 1b). Apart from a second major vc group (different from the EAN) which occurred at 10–16% in the Tewkesbury–Gloucestershire area, the remainder again consisted of mainly unique vc groups.

In addition, A-mating type isolates, which are typically of low frequency in aggressive populations, probably as a result of poorer pathogenic ability^{9–11}, occurred at ~10–40% in the

heterogeneous elements of these populations, yet were absent in the supergroups and second major vc groups. This raised the question of the status of the supergroups. Had they been solely the products of founder effects during epidemic spread, their subsequent replacement through frequent sexual recombination^{12,13} could have been expected. But their continued survival into the post-epidemic period, together with the absence of the A-mating type, suggested that they could also be favoured by selection. The higher frequency of the EAN supergroup in the Polish as compared with the Romanian sample (Table 1; $P < 0.01$) also suggested that the supergroups could be associated in particular with epidemic fronts.

In 1984 two new NAN epidemic fronts were located for further study near Madrid (Table 1c). At Avila, the population consisted largely of a single vc group, although it differed from the NAN supergroup. An unusual feature was the high frequency of A-mating types. The small remainder was heterogeneous. At Guadalajara the outbreak also consisted largely of a single vc group, the same as the second major group found at Tewkesbury (Table 1b), and the A-mating type was absent, as it was in Britain. These results therefore demonstrated that vc groups other than the EAN and NAN supergroups could be dominant, and supported the hypothesis of population uniformity at epidemic fronts.

In 1985 a third outbreak area was defined, centred on Lisbon. At fresh epidemic fronts to the east of Porto and to the north and south-east of Lisbon (Fig. 1), all 47 samples proved to be of the NAN supergroup, and again no A-types were found. At sites in older epidemic areas, of an estimated 3–6 years standing, and closer to the probable original outbreak centre in Lisbon, the supergroup was still the major component at ~60%, but the remainder comprised 25 vc groups among 40 isolates, and 15% were A-types. These data not only provided further evidence of the uniformity of epidemic front populations, but indicated that behind the fronts new vc groups were appearing in association with the A-mating type, presumably an intermedi-

ate step between a near clonal front and a mainly heterogeneous post-epidemic population.

All the above samples were from the pathogenic phase of the fungus in the xylem of diseased elm twigs. But the heterogeneity would be most likely to originate during the saprophytic phase of the fungus in diseased elm bark, where the sexual stage occurs^{9,12,13}. To investigate the origin and rate of appearance of the heterogeneity, two Lisbon area sites, one a frontal site at Tomar and the other an older site at Mafra (T and M, Fig. 1), were resampled in October 1986 by collecting pathogenic phase (twig) and saprophytic phase (bark) samples at each. At the epidemic front (Tomar), both the pathogenic and saprophytic phases now showed a small heterogeneous component, although with a high ratio of isolates/vc groups in it of 5.0 (Table 2). As expected, this component was much larger at the older site (Mafra); moreover, its isolate/vc group ratio was much lower at ~1.5. The heterogeneous component was also significantly larger in the saprophytic phase compared with the pathogenic phase at both sites ($P < 0.001$). Most importantly, ~90% of the emerging heterogeneous component at the front was of the normally scarce A-mating type, whereas this comprised only 20–40% at the older site ($P < 0.001$ for saprophytic phase, Tomar versus Mafra).

Evidently the observed heterogeneity can arise very rapidly, and is probably generated in the saprophytic phase. It would be difficult to determine whether the new vc groups arise entirely *de novo*, for example through mutation and pseudo-selfing^{14,15}, or are already present at very low levels; however, their rate and pattern of emergence is consistent with the former. It also coincides with a remarkable explosion of A-mating types (compare with Poland, Table 1) and the transient appearance of a small number of supergroup A-types in the pathogenic phase (Tomar, Table 2). As A-types are especially fertile^{9,10,14}, this is likely to result in a burst of sexual reproduction, much of which would, initially, involve the outcrossing of the novel vc group A-types with the dominant B-type supergroup component. The increasing heterogeneity in the saprophytic phase presumably feeds through to the pathogenic phase, as at Mafra (Table 2), until eventually a relatively stable population structure, such as that in Romania or Britain (Table 1), is reached consisting of a small all B-type supergroup component and a large highly heterogeneous component with a low frequency of A-mating types. Indeed, the Mafra population seems to be approaching this point.

Other evidence suggests that the change to vc heterogeneity may be associated with the d-factor, a debilitating virus-like

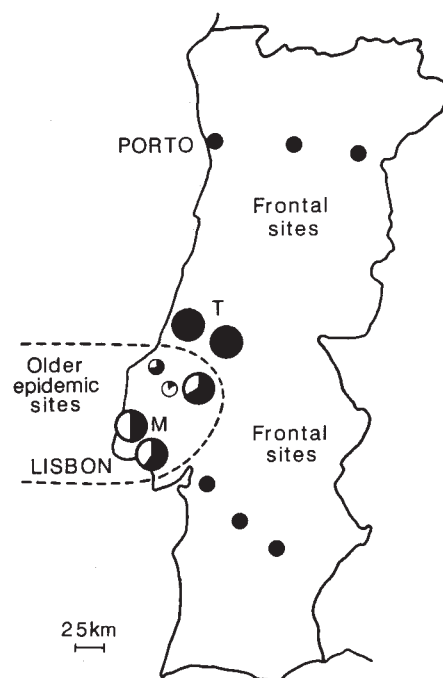


Fig. 1 Distribution of NAN supergroup (●) and heterogeneous component (○) genotypes in a survey of Portugal in 1985. Note predominance of the supergroup at epidemic front sites and presence of the heterogeneous component at older epidemic sites. Smaller circles, 2–14 samples; larger circles, 22–38 samples. The non-aggressive subgroup (not shown) represented ~74% of the overall *O. ulmi* sample at the fronts and ~5% at the older epidemic sites. T, Tomar; M, Mafra (see Table 2).

cytoplasmic disease of *O. ulmi*^{16–20}. This spreads readily by hyphal fusions between colonies of the same vc group and into asexual conidiospores, but is strongly restricted by vegetative incompatibility reactions and is not transmitted into sexually produced ascospores^{16–19}. Hence dominant vc groups are likely to be at high risk from spread of d-factors, and some of the 'frontal' vc clones do show very high levels of d-infection: ~31% in Poland and ~49% in Guadalajara (Table 1a and c). In contrast, comparable post-epidemic supergroup samples from Romania and Britain show only sporadic d-infection (0–18%;

Table 2 Comparison of current epidemic and recent epidemic NAN aggressive samples from Portugal, 1986

Site	Status	NAN sample size*	Uniform or supergroup component		Heterogeneous component					
					vc group frequency					
			% of sample	% A type	% d- infected	% of sample	No. of groups/no. of isolates	Ratio isolates/ groups	% A type	% d- infected
Tomar	Current epidemic front									
	Pathogenic phase	59	93	4	13	7	(3/4)	(1.33)	(75)	(0)
	Saprophytic phase	150	64	5	44	36	9/45†	5.0	91	38
Mafra	Recent epidemic front									
	Pathogenic phase	57	37	0	0	67	21/29	1.38	19	6
	Saprophytic phase	102	9	33	67	91	52/83	1.60	46	21

* The non-aggressive subgroup *O. ulmi* still comprised ~21% of the overall Tomar twig sample and ~41% of the overall bark sample, but was absent in the Mafra samples.

† One vc group of 24 isolates, comprising 20 A- and 4 B-types, plus two vc groups of 5 isolates, two of 3, one of 2 and three of 1 isolate, all A-types.

Twig or pathogenic phase isolates were sampled as in Table 1. For bark or saprophytic phase samples, slabs of inner bark containing active beetle-breeding galleries were collected from mature dying elms. Isolations were made on to a selective medium²⁷ from the inner bark surface on a grid basis^{10,13} at 2–3 cm intervals. An area of 1,200 cm² bark was sampled for Tomar (two trees) and 600 cm² for Mafra. d-Infection levels were estimated on the basis of diseased colony phenotype^{16,17} after growth on malt extract agar²⁷ at 20 °C for 10 d in a single replicated test involving all isolates, and ability to give a d-reaction (see Table 1) in the case of supergroup isolates. Other methods as described in Table 1 legend.

Table 1a and b). In the Portuguese samples (Table 2), high d-infection levels were not recorded in the pathogenic phase, but in the saprophytic phase d-infection of the supergroup increased from ~44% at the front (Tomar) to ~67% (small sample) at the older site, whereas that of the heterogeneous component declined from ~38% to ~21%. As sexual outcrossing would tend to restrict the spread of d-factors by generating new vc groups and producing d-factor free ascospores, debilitation of the frontal vc clones by high levels of mycovirus infection could be a prime cause of the rapid change in population structure. The contrasting low level of d-infection in the dominant vc group at Avila (Table 1c) could therefore be due to the high frequency of A-mating types within it.

The Tomar/Mafra samples also support the view that the supergroups are products of selection maintained by asexual propagation. This is indicated primarily by their having a very uniform colony phenotype, whereas the heterogeneous component isolates are extremely variable. Second, in a pathogenicity test on a 4 m Commelin elm in June 1987, ten supergroup isolates (5 T + 5 M) were significantly more pathogenic than ten heterogeneous component isolates, causing 69 and 49% mean defoliation at 12 weeks, respectively ($P < 0.001$). Third, strong selection between the saprophytic and pathogenic phases is indicated (Table 2) by: (1) the smaller heterogeneous component in the pathogenic phase at both sites ($P < 0.001$); (2) the lower frequency of less pathogenic A-types^{9,11} in the pathogenic phase at Mafra ($P < 0.001$); (3) the lower frequency of the weakly pathogenic non-aggressive subgroup^{6,9} in the pathogenic phase at Tomar ($P < 0.01$) and (4) the lower frequency of diseased or d-infected isolates in the pathogenic phase at both sites ($P < 0.001$ for supergroup; $P < 0.01$ for heterogeneous component). The supergroup B types may therefore have a selective advantage over many heterogeneous component genotypes in the pathogenic phase, at least during the early stages of an outbreak. This in turn would enhance the supergroup genotype's success by increasing its feedback^{9,21} from the pathogenic phase into the saprophytic, so promoting its dispersal by the next generation of vector beetles. The population frequency of a supergroup may therefore depend in part upon a balance between its fitness advantage and the disadvantage of heavy d-infection. Other environmental changes as an epidemic progresses, such as the ensuing collapse of the elm and beetle vector populations^{5,9}, may also favour an increase in pathogen diversity and so contribute to the decline of the supergroups.

Further research is in progress on the comparative fitness and genetic structure of the dominant vc clones, on the selective impact of d-factors and on the processes underlying the appearance of A-mating types and rapid generation of variability. The latter could involve pseudo-selfing^{14,15} or gene-flow from the resident, but almost totally genetically isolated^{5,9}, non-aggressive subgroup. Although the emergence of 'pathotype' clones in biotrophic pathogens of arable crops is well known^{22,23}, the plasticity shown by *O. ulmi* suggests a potential for rapid evolution under intense selection, and for the emergence of fitted clones in more ecologically complex necrotrophic pathogens of perennial hosts. Such selection might occur during epidemic spread in a geographically new and susceptible host population as here, during exposure to widespread crop monoculture, or other environmental pressure. Nonetheless, a clone could also be particularly vulnerable to further changes in selection, including the spread of a mycovirus. Indeed, there is evidence to suggest that both the decline of the chestnut blight epidemic in Europe during the 1950s (refs 24–26), and the decline of the first epidemic of Dutch elm disease in the 1920s to 1940s (A. G. Mitchell and C.M.B., unpublished results) could have followed the spread of mycoviruses in dominant vc clones. In the current epidemics of Dutch elm disease, however, the diversification of the *O. ulmi* supergroups, if a response to d-factor pressure, may have occurred too rapidly for the latter to limit the initial impact of the disease.

I thank Drs D. Serrao Nogueira and L. Cid-Ferreira (Lisbon), Dr F. Robredo and Marie Carmen-Munoz (Madrid), Dr N. Nanu (Timisoara, Romania), Dr S. Lukomski and K. Manka (Warsaw and Poznan) and others for assistance with field-sampling, Susan Kirk (FC) for technical assistance and the FC, EEC and British Council for financing field work abroad.

Received 30 November 1987; accepted 17 February 1988.

- Gibbs, J. N., Houston, D. R. & Smalley, E. B. *Phytopathology* **69**, 1215–1219 (1979).
- Houston, D. R. *Pl. Dis.* **69**, 677–680 (1985).
- Brasier, C. M. & Gibbs, J. N. *Nature* **242**, 607–609 (1973).
- Brasier, C. M. *Nature* **281**, 78–80 (1979).
- Brasier, C. M. in *Populations of Plant Pathogens* (eds Wolfe, M. S. & Caten, C. E.) 213–226 (Blackwell, Oxford, 1987).
- Gibbs, J. N. & Brasier, C. M. *Nature* **241**, 381–383 (1973).
- Peace, T. R. *For. Comm. Bull.* **33**, 1–44 (HMSO, London, 1960).
- Gibbs, J. N. *A. Rev. Phytopathol.* **16**, 287–307 (1978).
- Brasier, C. M. *Adv. Pl. Pathol.* **5**, 53–118 (1986).
- Brasier, C. M. in *The Ecology and Physiology of the Fungal Mycelium* (eds Jennings, D. H. & Rayner, A. D. M.) 451–497 (Cambridge University Press, 1984).
- Brasier, C. M. & Webber, J. F. *Pl. Pathol.* **36**, 462–466 (1987).
- Lea, J. & Brasier, C. M. *Trans. Br. mycol. Soc.* **80**, 381–387 (1983).
- Webber, J. F., Brasier, C. M. & Mitchell, A. G. in *Plant Infecting Fungi* (eds Pegg, G. & Ayres, P. D.) 298–313 (Cambridge University Press, in the press).
- Brasier, C. M. & Gibbs, J. N. *Nature* **257**, 128–131 (1975).
- Brasier, C. M. *Adv. Pl. Pathol.* **6** (in the press).
- Brasier, C. M. *Nature* **305**, 220–223 (1983).
- Brasier, C. M. in *Fungal Virology* (ed. Buck, K. W.) 177–208 (CRC, Boca Raton, Florida, 1986).
- Rogers, H. J., Buck, K. W. & Brasier, C. M. *Pl. Pathol.* **35**, 277–287 (1986).
- Rogers, H. J., Buck, K. W. & Brasier, C. M. in *Viruses of Fungi and Lower Eukaryotes* (eds Koltin, Y. & Leibowitz, M.) (Dekker, New York, in the press).
- Webber, J. F. *Pl. Pathol.* **36**, 531–538 (1987).
- Webber, J. F. & Brasier, C. M. in *Invertebrate-microbial Interactions* (eds Anderson, J. M., Rayner, A. D. M. & Watton, D.) 271–306 (Cambridge University Press, 1984).
- Wolfe, M. S. & Caten, C. E. (eds) *Populations of Plant Pathogens* (Blackwell, Oxford, 1987).
- Burdon, J. J. & Roelfs, A. P. *Phytopathology* **75**, 907–913 (1985).
- Grete, J. & Sauret, S. *C. r. Acad. Sci. Ser. D* **268**, 2347–2350 (1969).
- Anagnostakis, S. A., Hau, B. & Kranz, J. *Pl. Dis.* **70**, 536–538 (1986).
- Anagnostakis, S. A. *Mycologia* **79**, 23–37 (1987).
- Brasier, C. M. in *Compendium of elm diseases* (eds Stipes, R. J. & Campana, R. J.) 76 (American Phytopathological Society, St Paul, Minnesota, 1982).

Molecular basis for evasion of plant host defence in bacterial spot disease of pepper

Brian Kearney, Pamela C. Ronald,
Douglas Dahlbeck & Brian J. Staskawicz

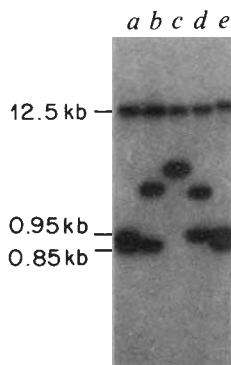
Department of Plant Pathology, University of California, Berkeley, California 94720, USA

A fundamental objective in studying host-pathogen interactions is to determine how a pathogen evolves to overcome the defences of a previously resistant host. In mammalian systems, evolution of a pathogen population is often directed towards avoiding recognition by an immune system^{1–3}. Although plants lack circulating antibodies and the memory response of mammalian immune systems, it is becoming increasingly clear that active resistance in plants involves recognition of the pathogen by its host^{4,5}. In this paper we present the first molecular evidence to indicate that plant pathogens evolve to overcome resistance by evading host recognition and response. The bacterial pathogen *Xanthomonas campestris* pathovar *vesicatoria* mutates to overcome genetically defined resistance in pepper, *Capsicum annuum*, by the transposon-induced mutation of *avrBs1*, a bacterial gene that provokes the plant's resistance response.

Bacterial spot disease of peppers is caused by *X. c. pv. vesicatoria* (*Xcv*)^{6–8}. Peppers carrying the gene *Bs1* are resistant to Race 2 strains of *Xcv*. Inoculation of the cultivar ECW10R (*Bs1Bs1*) with *Xcv* Race 2 causes a hypersensitive response⁹ (HR) in the region of infection. When ECW10R is inoculated

Fig. 1 Southern blot of *Xcv* strains probed with *avrBs₁*. Lane a, DNA from *Xcv* Race 2, strain 81-23. Three hybridizing bands appear at 12.5, 0.95 and 0.85 kb. Lanes b–e, DNA from M1, M2, M4 and M13 respectively. Each mutant displays a mobility shift in one hybridizing band of ~1.2 kb when compared with the wild-type isolate. In M2 both the 0.85 kb and 0.95 kb bands disappear and a new band appears at 3.0 kb, which can be explained by a 1.2 kb insertion into the *EcoRV* site in between the two smaller bands.

Methods. Total genomic DNA was digested with *EcoRV*, separated electrophoretically through an 0.8% agarose gel, transferred to Nytran membrane and probed with a ³²P-labelled *Pst*I/*Bgl*III fragment of pXV2007¹² containing *avrBs₁*. The blots were washed at high stringency, essentially as described¹⁹.



with *Xcv* Race 1 water-soaked lesions occur, and the pathogen is able to multiply to high titre¹⁰.

To study how bacterial pathogens become virulent on previously resistant hosts, thirteen independent spontaneous mutants of *Xcv* Race 2 were selected that had become virulent on ECW10R and were thus able to overcome *Bs₁*^{11,12}. We have previously identified and cloned a genetic locus (*avrBs₁*) from *Xcv* Race 2 that is responsible for specifically inducing a HR on pepper plants carrying the resistance gene *Bs₁*. When a plasmid containing *avrBs₁* was conjugated into each of the 13 spontaneous mutants, the transconjugants caused a HR on pepper plants containing *Bs₁*¹². We therefore concentrated our study on the *avrBs₁* locus of each mutant.

Four mutants were chosen for detailed study. Of the four, two (M1 and M13) had completely overcome the resistance encoded by *Bs₁* and induced fully water-soaked lesions when inoculated onto ECW10R plants. The other two mutants caused intermediate reactions on *Bs₁* plants. M4 develops a HR on ECW10R ~24 h later than the wild-type pathogen. M2 causes

an intermediate reaction that is partially necrotic and partially water-soaked.

The *avrBs₁* locus of each mutant was examined and compared to the wild-type gene from *Xcv* Race 2, strain 81-23. Southern analysis of *EcoRV*-digested DNA revealed that ~1.2 kilobases (kb) of DNA was inserted into different regions of *avrBs₁* in each mutant (Fig. 1). Southern analysis of the remaining nine mutants demonstrated that 1.2 kb of DNA was inserted into *avrBs₁* in these cases as well (data not shown).

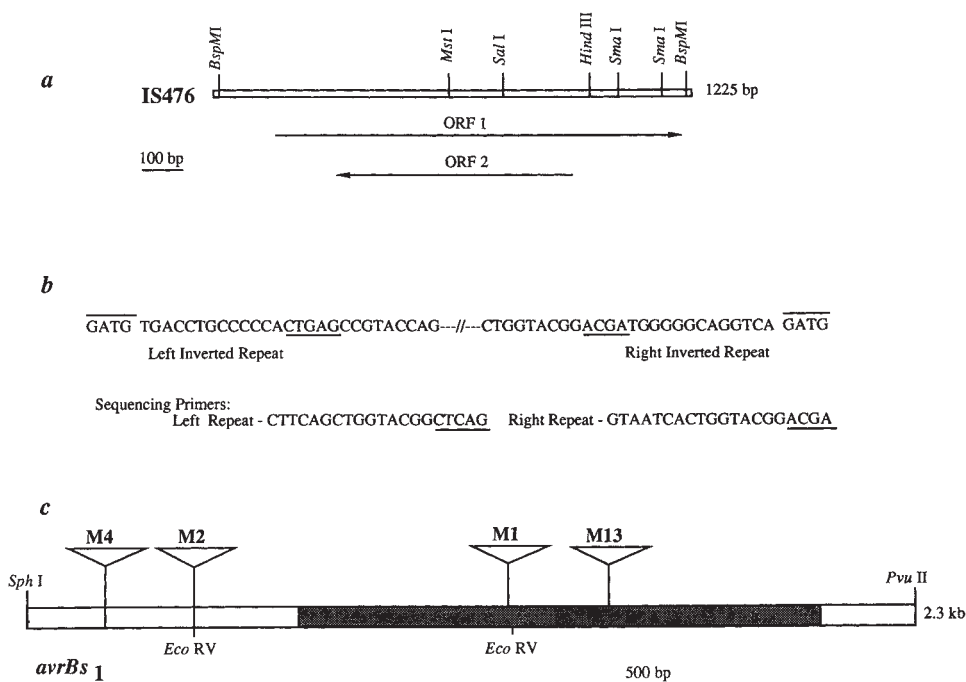
To further analyse the mutant avirulence genes, a library of *Xcv* Race 2, mutant M1, was made in the cosmid pLAFR3 essentially as described previously¹³. Two clones containing the mutant *avrBs₁* gene were identified by colony hybridization using a 5.3 kb *Pst*I/*Bgl*III fragment containing *avrBs₁* as a probe to screen the library. To clone the *avrBs₁* locus from the other three mutants under detailed study, total genomic DNA was digested with *Pst*I and *Sst*I, size fractionated on an agarose gel, and 6–9 kb fragments were cloned into pUC118 and pUC119¹⁴. The clones containing the mutant *avrBs₁* loci from M2, M4 and M13 were also identified using the wild-type 5.3 kb fragment containing *avrBs₁* as a ³²P-labelled probe.

Restriction-enzyme analysis of the cloned loci confirmed that each mutation was accompanied by an insertion of 1.2 kb into *avrBs₁*. Furthermore, the DNA inserted in each case had an identical restriction map (Fig. 2a), suggesting the independent mutations were caused by the same element.

The complete DNA sequence of the insertion from M1 was determined. The element is 1,225 base pairs (bp) long and displays the structures of a typical bacterial insertion sequence. It has 26 bp imperfect inverted repeats and is flanked by a 4 bp target-site duplication. Two open reading frames (1,037 and 593 bp) are present in opposite orientations in the same frame. The element has been listed with the Central Plasmid Registry¹⁵ and has been named IS476.

The precise location of the insertion in each mutant was determined by sequencing outward into *avrBs₁* from the ends of IS476. A 20-nucleotide primer was synthesized for each inverted repeat. Small differences in the sequences of the left and right repeats were exploited to make specific 3' ends of the primers (Fig. 2b). The location of IS476 in the mutants is shown

Fig. 2 a, Restriction enzyme map of IS476. For all the sites tested, the restriction map of IS476 in mutants M1, M2, M4 and M13 is identical. b, The 26 bp inverted repeats and the target site duplication (GATG) from mutant M1 are shown. The regions underlined indicate the differences between the left and the right repeats. The sequencing primers synthesized to localize the position of insertion of IS476 in each mutant are also shown. The underlined bases at the 3' ends of each primer correspond to the areas of difference between the inverted repeats. c, Restriction map of *avrBs₁* showing the position of insertion of IS476 in each mutant as determined by sequencing analysis. The shaded region of the map contains the open reading frame necessary for *avrBs₁* activity (P.C.R. *et al.*, manuscript in preparation). In M1 and M13 IS476 inserted into the coding region of *avrBs₁*. In M4 and M2, IS476 inserted far upstream of the *avrBs₁* coding region, suggesting that the mutations are polar or lie in regulatory regions necessary for normal *avrBs₁* activity.



in Fig. 2c. In M1 and M13 (which completely lose *avrBs₁* function), IS476 is integrated into the coding region of *avrBs₁* (shaded area in Fig. 2c; P.C.R. *et al.*, manuscript in preparation). In both M4 and M2, IS476 is inserted upstream of the *avrBs₁* coding region. As mentioned previously, the HR caused by M4 on *Bs₁* plants is delayed by 24 hours and M2 gives an intermediate HR/water-soaking phenotype on *Bs₁* plants. These intermediate reactions may result from lowered protein levels due to IS476 insertion into an *avrBs₁* regulatory region.

In bacterial spot disease of pepper, the stability of resistance is a function of the mutation frequency of the avirulence genes in the pathogen, and, to date, all naturally occurring mutations isolated in *avrBs₁* have been caused by IS476. We do not know if transposon mutagenesis is a universal mechanism by which pathogens overcome host resistance as too few data are available to make any conclusion. Presumably any kind of mutation that enables host recognition to be avoided would have the same effect.

The importance of IS476 in bacterial spot disease may result from its linkage to copper resistance in *Xanthomonas*. Southern analysis has shown that the 200 kb plasmid (pXvCu1) (ref. 16) in *Xcv* that carries *avrBs₁* also carries three copies of IS476 and a gene for copper resistance. We have recently determined that at least one copy of IS476 on pXvCu1 is an active transposable element, and preliminary data suggest that mutation rates in *Xcv* backgrounds containing IS476 are approximately an order of magnitude higher than in backgrounds lacking the element (B.K. *et al.*, manuscript in preparation). We have also determined that pXvCu1 is self-transmissible in bacteria grown both in culture^{12,16} and in plants when maintained under copper selection (data not shown). Copper has been a primary chemical defence against bacterial spot disease in Florida for years^{17,18}, and now almost all strains of *Xcv* isolated from Florida are copper resistant (R. E. Stall, personal communication). Southern blot analysis of DNA isolated from a 20-year collection of *Xcv* strains has shown that IS476 is found only in copper-resistant strains (data not shown). This evidence suggests that IS476 is spread by the copper-selected transmission of pXvCu1. It is ironic that by trying to control *Xcv* pathogenesis with copper sprays, a highly mutagenic element has been propagated that enables the pathogen to overcome genetically defined disease resistance.

We thank Robert Stall for the *Xcv* strains and pepper hosts, Jean Swanson for isolating M1 and initiating its analysis, and Milt Schroth, Thanh Hyunh, Ulla Bonas, Maureen Whalen, and Susan Latshaw for critical reading of the manuscript and helpful suggestions. This work was funded by NSF. B.K. was supported by a NSF Graduate Fellowship.

Received 14 December 1987; accepted 3 March 1988.

1. Palese, P., & Young, J. F. in *Genetics of Influenza Viruses* (eds Palese, P. & Kingsbury, D. W. 321-336 (Springer, Vienna, 1983).
2. Silverman, M., & Simon, M. in *Mobile Genetic Elements* (ed. Shapiro, J. A.) 537-557 (Academic, New York, 1983).
3. Van der Ploeg, L. H. T. *Cell* **51**, 159-161 (1987).
4. Keen, N. in *Advances in Plant Pathology* Vol. 1 (eds Ingram, D. S. & Williams, P. H.) 35-83 (Academic, New York, 1982).
5. Ralton, J. E., Howlett, B. J. & Clarke, A. E. in *Intercellular and Intracellular Communication I. Hormones, Receptors and Cellular Interactions in Plants* (eds Chadwick, C. M. & Garrod, D. R.) 281-318 (Cambridge University Press, New York, 1986).
6. Cox, R. S., Conover, R. A. & Sowell, G. *Phytopathology* **46**, 582-589 (1956).
7. Diab, S., Bashan, Y., Okom, Y. & Henis, Y. *Phytopathology* **72**, 1257-1260 (1982).
8. Kennedy, B. W. & Alcorn, S. M. *Pl. Dis.* **64**, 674-676 (1980).
9. Klement, Z. in *Phytopathogenic Prokaryotes* Vol. 2 (eds Mount, M. S. & Lacy, G. H.) 149-177 (Academic, New York, 1982).
10. Cook, A. A. & Stall, R. E. *Pl. Dis.* **66**, 388-389 (1982).
11. Dahlbeck, D. & Stall, R. E. *Phytopathology* **69**, 634-636 (1979).
12. Swanson, J., Kearney, B., Dahlbeck, D. & Staskawicz, B. *Molec. Pl. Microbe Interact.* **1**, 5-9 (1988).
13. Staskawicz, B., Dahlbeck, D., Keen, N. & Napoli, C. J. *Bacteriol.* **169**, 5789-5794 (1987).
14. Vieira, J. & Messing, J. *Meth. Enzym.* **153**, 3-11 (1987).
15. Campbell, A., Berg, D. E., Novick, R. P., Starlinger, P. & Szybaski, W. *Plasmid* **2**, 466-473 (1979).
16. Stall, R. E., Loschke, D. C. & Jones, J. B. *Phytopathology* **76**, 240-243 (1986).
17. Marco, G. M. & Stall, R. E. *Pl. Dis.* **67**, 779-781 (1983).
18. Cox, R. S. *Pl. Dis.* **66**, 870 (1982).
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

DNA typing from single hairs

Russell Higuchi, Cecilia H. von Beroldingen*, George F. Sensabaugh* & Henry A. Erlich

Department of Human Genetics, Cetus Corporation, Emeryville, California 94608, USA

* Forensic Sciences Program, University of California, Berkeley, California 94720, USA

The characterization of genetic variation at the DNA level has generated significant advances in gene and disease mapping¹, and in the forensic identification of individuals²⁻⁶. The most common method of DNA analysis, that of restriction fragment length polymorphism (RFLP), requires microgram amounts of relatively undegraded DNA for multi-locus typing, and hundreds of nanograms for single-locus comparisons⁷. Such DNA frequently cannot be obtained from forensic samples such as single hairs and blood stains, or from anthropological, genetic or zoological samples collected in the field. To detect polymorphic DNA sequences from single human hairs, we have used the polymerase chain reaction (PCR), in which specific short regions of a gene can be greatly amplified *in vitro*⁸⁻¹⁰ from as little as a single molecule of DNA¹⁰. We have detected genetically variable mitochondrial and nuclear DNA sequences from the root region of shed, as well as freshly-plucked, single hairs; mitochondrial DNA (mtDNA) sequences have been detected in a sample from a single hair shaft. We have used three different means of DNA typing on these samples: the determination of amplified DNA fragment length differences, hybridization with allele-specific oligonucleotide probes, and direct DNA sequencing.

Most of the DNA in hair is located in the root and surrounding sheath cells¹¹. As measured by the fluorescence of DNA-dye complexes in crude lysates of hairs, the root end of freshly-plucked hairs may contain as much as 0.5 µg DNA, whereas hair shafts contain too little DNA to determine its quantity and condition¹¹. Our recovery of purified DNA, however, has rarely been more than 200 ng from freshly-plucked hairs and has usually been less than 10 ng from shed hairs. DNA 'fingerprint' analysis with multi-locus probes has been done on DNA pooled from multiple, freshly-plucked hairs of the same individual³ and a single-locus analysis could presumably be performed on a single, freshly-plucked hair⁷. The hairs found at the scene of a crime, however, may derive from different individuals, and are usually shed hairs. Therefore, the typing of single, shed hairs would be much preferable to that of pooled, freshly-plucked samples.

Figure 1 shows electrophoresis of the amplification products from DNA isolated from the separated root and shaft portions of single, freshly-plucked head hairs from two individuals. The Klenow fragment of DNA polymerase I was used for these amplifications^{8,9}. A previously characterized length polymorphism in human mtDNA^{12,13} distinguishes the samples from the different individuals; DNA sequencing of the mtDNA of individual 1 had shown a 9 base-pair (bp) deletion in this region¹³. This deletion may be Asian-specific. PCR with oligonucleotide primers flanking this region results in a DNA fragment that includes the length variable region and is either 111 or 120 bp depending on whether this deletion is present (this difference could, therefore, be deemed a '(PCR)FLP'). Amplification products of the appropriate sizes for each individual are obtained from both the root and shaft portions of the hairs. Thus even the shaft portion contains enough copies of mtDNA to be detected using PCR. Electrophoresis can also distinguish amplification products that differ in their pattern of restriction enzyme sites^{3,9,14}.

Figure 2 shows typings of a nuclear gene amplified from the root portions of single, shed hairs from six different individuals. 'Single-copy' nuclear genes, in this case the class II HLA gene

Fig. 1 PCR amplifications of mtDNA from single hairs. Hairs from two individuals were separated into 'root' (containing bulb, sheath, elongation zone and some shaft¹¹) and shaft portions. DNA was prepared and PCR amplification performed as below. Fragment lengths were determined by electrophoresis. Lane 1, PCR product from the root portion of a head hair of individual 1; lane 2, PCR product from the shaft portion of the same hair; lane 3, PCR product from the root portion of a head hair of individual 2; lane 4, PCR product from the shaft portion of the same hair. **Methods.** Single hairs ~10 cm in length were plucked, cut with scissors and rinsed in distilled H₂O followed by absolute ethanol. After drying, the hair-portion was digested as in ref. 3, in 0.5 ml of 0.01 M Tris-HCl, 0.01 M EDTA, 0.1 M NaCl (pH 8.0) containing 50 µg ml⁻¹ proteinase K, 0.039 M DTT, and 2% SDS. DNA was extracted with equal volumes of phenol/chloroform followed by *n*-butanol. The aqueous phase was added to 2 ml of TE (0.01 M Tris-HCl, 0.1 mM EDTA (pH 8.0)) in the upper portion of a Centricon 30 ultrafiltration device (Amicon). The sample was centrifuged according to manufacturer's instructions, the retained portion resuspended in 2 ml of TE and recentrifuged. The desalted DNA was recovered in ~40 µl. PCR amplification using the Klenow fragment of DNA polymerase I was carried out on 15 µl of this solution in a final volume of 100 µl which included the reagents described in ref. 9 along with primers specific to an intergenic region of human mtDNA¹³. Cycles (30) of PCR were performed as described^{8,9}. DNA from half of the amplification reactions was subjected to electrophoresis on composite 3%, 'Nu-sieve' agarose (FMC)/1% non-modified agarose gels. The gel was stained with ethidium bromide and photographed under UV light, as shown.

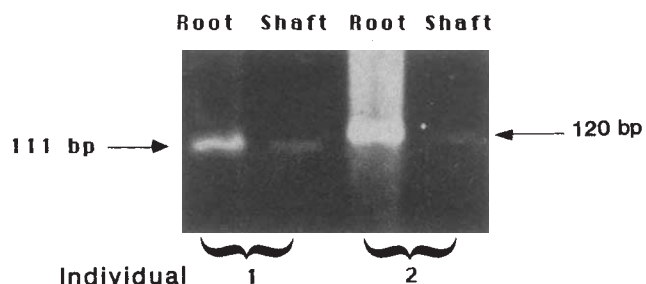
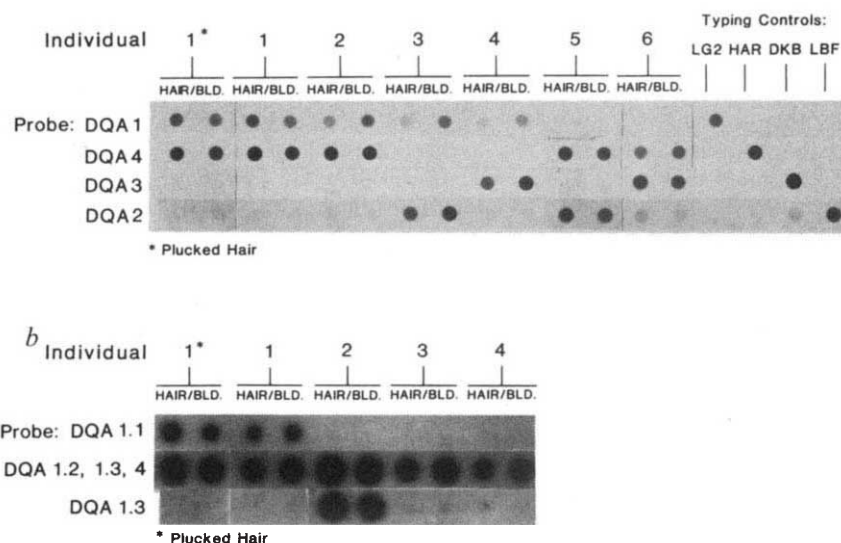


Fig. 2 HLA typing on single shed hairs. (a) The samples are from either the root portion of a single, shed hair (Hair) or from DNA prepared from fresh blood of the same individual (Bld.). A freshly plucked hair from individual 1 is also included. For each sample, PCR product amplified from exon 2 of the HLA class II locus, *DQA*, was fixed on nylon membranes in a pattern of four dots per sample. Each dot was tested for hybridization with an oligonucleotide probe conjugated with horseradish peroxidase (HRP; C. Cheng, manuscript in preparation) and specific for one of four alleles defined by sequences between codons 47 and 56 of this gene¹⁶ (note: nomenclature for these sequences has been revised from that in ref. 16 to that in ref. 29). HRP was detected by its localized catalysis of the conversion of the colourless substrate, tetramethylbenzidine, to a blue precipitate³⁰; the retention of an HRP-probe corresponding in sequence to a given allele indicates the presence of that allele in the sample. The rightmost four samples are control PCR products amplified from the DNA of cell lines homozygous for the indicated *DQA* types as defined by DNA sequence. They are: LG2 (serological type-DR1, DQw1); HAR (DR3, DQw2); DKB (DR9, DQw3) and LBF (DR7, DQw2). Typing results are: individual 1-(DQA 1, 4); individual 2-(1, 4); individual 3-(1, 2); individual 4-(1, 3); individual 5-(2, 4); individual 6 (3, 4). Although weak spots appear in several places on the membrane exposed to probe DQA2, comparison with the typing controls, for example DQA2 versus DKB, shows them not to be elevated over background. (b), Sub-typing of samples possessing the DQA1 allele. 'Dot-blot' of the same amplifications from individuals 1-4 used above were probed with allele-specific oligonucleotides derived from the codon 47-56 region of *DQA*. Note that the probe DQA1.2 cross-hybridizes with alleles 1.3 and 4. Therefore, an additional probe is required to distinguish the 1.2, 1.3 from the 1.3, 1.3 genotype. These probes were labelled with ³²P and their retention detected by autoradiography. Combining these results with those above, individuals 1-4 type as (1.1, 4); (1.3, 4); (1.2, 2); and (1.2, 3), respectively.



Methods. Individual DNAs from were prepared from the root ends of hairs as in Fig. 1. PCR amplifications were done, however, with the thermostable DNA polymerase of *T. aquaticus* (D. Gelfand *et al.*, unpublished data and refs 10, 31). PCR was on 15 µl aliquots of the DNA from each hair in 100 µl volumes of 0.01 M Tris-HCl (pH 8.3), 0.05 M KCl, 2.5 mM MgCl₂, 0.188 mM of each dNTP containing 100 µg ml⁻¹ gelatin, 20 pmol of each primer¹⁶ and 2.5 units of *Taq* DNA polymerase. Cycles (40) of PCR took place in an automated temperature cycling device (Perkin-Elmer Cetus) that provided, per cycle, 1 min each of incubation at 95, 55, and 70 °C. In a dot-blot manifold (S & S), ~50 ng of specific PCR product was immobilized in each dot to a Genatrans nylon membrane (Plasco) which was then exposed to UV light to fix the DNA³². In (a), membranes were hybridized as in R. Saiki *et al.* (manuscript submitted), with HRP-probes at a concentration of 1 pmol ml⁻¹. The final wash was at 40 °C in 0.1 × SSPE, 0.1% Triton X-100. Colour development was as described in ref. 30. In (b), membranes were hybridized as in ref. 16 except that the final wash before autoradiography was at 42 °C in 0.1 × SSPE, 0.1% SDS.

DQA, are a greater challenge to detect than are mtDNA sequences, which are present in many hundreds of copies per cell. Nuclear genes, however, offer greater potential variation and, therefore, more informative typing. In this experiment, the higher efficiency of PCR using a thermostable DNA polymerase¹⁰ facilitated the amplification and typing of a 242 bp region of the *DQA* gene¹⁶. Alleles of this gene can be distinguished using allele-specific oligonucleotide (ASO) probes against immobilized 'dots' (dot-blot) of the amplified DNA. The typing in panel A was done using horseradish peroxidase-coupled, ASO probes (R. Saiki, *et al.*, in preparation). Such a

non-radioactive format, made possible by the great amplification of target sequences, would benefit forensic and other laboratories unable to use radioisotopes. The ASO probes are specific for the four alleles defined by sequences between codons 47 and 56 of *DQA*¹⁶. The allele defined by the DQA1 probe can be further subdivided, as shown in panel b, by three additional ASO probes to the region between codons 30 and 45 (ref. 16). The allele frequencies defined by these probes are shown in Table 1. The distribution of these alleles and the twenty-one genotypes they define in Caucasian, black and Asian populations will be the subject of a subsequent report (G. Horn and

Table 1 Frequency of alleles of *DQA*

Allele	1	2	3	4
Frequency	0.380	0.105	0.212	0.302
Subtyping of allele 1:				
Allele	1.1	1.2	1.3	
Frequency	0.139	0.207	0.034	

Individuals were of Caucasian, Black and Asian ancestry. For each of these populations, the frequency of genotypes are in approximate Hardy-Weinberg equilibrium. Number of alleles sampled, 410.

R. Helmuth, unpublished data). The observed distributions of genotypes within each population are consistent with the expected frequencies based on Hardy-Weinberg equilibrium.

As shown, the *DQA* genotypes of these shed hairs matched in all cases those found in the DNA from the nucleated cells of peripheral blood of the same individual. From shed hairs we usually recover less than 10 ng DNA, that is ~1,000 copies of this HLA gene. Amplifications of the *DQA* gene have been successful on single copies of this gene (U. Gyllensten, personal communication), as have been amplifications of single copies of the human β -globin gene¹⁰. In subsequent experiments we have had success typing from several-month-old, fallen hairs in which we fail to detect DNA by chemical means (<1 ng DNA). Thus far we have typed nearly a hundred individual hairs using the ASO system for *DQA* on DNA amplified by PCR.

Whereas Figs 1 and 2 show DNA typings based on nucleic acid sequence differences, Fig. 3 shows the polymorphic DNA sequences themselves, determined rapidly and directly from single hair roots. In this case, a part of the D-loop region of human mtDNA was amplified. DNA sequencing, without cloning, was performed directly on the amplified DNA^{13,17}. Extensive sequence polymorphism exists in the D-loop (ref. 18 and P. Bluford *et al.*, unpublished results). As seen in Fig. 3, the two individuals are distinguished by these DNA sequences. The sequence from each individual's hair matches that obtained from DNA from the same individual's blood (data not shown).

The use of a thermostable DNA polymerase has made it possible to automate PCR¹⁰. Recent advances in automating DNA sequencing^{19,20} suggest that these could be combined to provide an automated DNA typing based on DNA sequence or fragment length determinations. Given that sample preparation for PCR could be simplified¹⁶ and automated, a machine capable of determining informative DNA sequences from a single hair is a very real possibility.

Hairs are one of the most frequently found forms of biological evidence at crime scenes^{21,22}, so their identification can be of considerable forensic importance. The current identification of hairs is based mainly on morphology and has an inherent subjectivity^{21,22}, although protein polymorphisms have been detected in, and ABH grouping performed on single hairs²²⁻²⁴. Protein polymorphisms in hair do not have the great potential for individual identification that DNA typing offers. It should be possible to simultaneously amplify more than one gene, for example, mtDNA as well as HLA genes, from a single sample so that polymorphic systems can be combined to give a high degree of individual specificity. The ability to determine genotype directly promises to provide the forensic hair specialist with objective criteria for individual identification.

For the medical anthropologist the collection of hair samples, rather than blood, may also be useful in those circumstances where the giving of blood is forbidden by religion or custom and/or where transport of blood samples may be difficult or expensive. For the zoologist, the collection of shed or even plucked hair may be preferable to the collection of a blood sample (O. Ryder, personal communication) which may be difficult (for example, from an elephant or from a large carnivore), or ill-advised (for example, from a member of an endangered species that should not be disturbed).

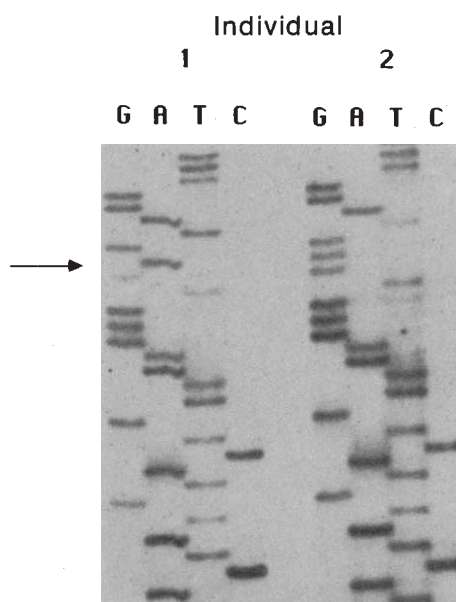


Fig. 3 Mitochondrial DNA sequence determined directly from a single hair root. A 228 bp portion of the D-loop region of human mtDNA was amplified from DNA obtained from single hair roots using as primers oligonucleotides corresponding to the human mtDNA positions³³ 16,192 to 16,209 (primer D18) and to the complement of the sequence from 16,401 to 16,420 (primer D6). Preparation of hair DNA and PCR was as in Fig. 2. Sequencing of amplified DNA^{13,17} used the dideoxy-termination method³⁴ and the modified T7 DNA polymerase, Sequenase (US biochemical). Shown are portions of autoradiographs of sequencing gels of amplified D-loop DNA from two individuals. The arrow points to a single nucleotide difference between them. DNA was also isolated and D-loop DNA amplified and sequenced from blood of the same individuals. In both cases, the sequences determined from blood and hair of the same individual were the same (data not shown). **Methods.** The amplified DNA was purified by electrophoresis on a 4% Nusieve (FMC) agarose gel. The predominant DNA band was cut out of the gel and placed in 0.1 ml of TE (0.01 M Tris-HCl, pH 8.0, 0.1 mM EDTA). The gel slice in TE was frozen and thawed twice to help elute DNA³⁵; a 20 μ l portion of the supernatant was added to a new 0.1 ml PCR reaction with the same primers as above (20 pmol each). Additional PCR cycles (15) were performed. This gel-elution/re-amplification yields ~1 μ g of predominantly D-loop sequence. The re-amplification product was treated as in ref. 13 except that the PCR reaction was extracted once with an equal volume of 1:1, phenol/CHCl₃ (equilibrated with TE), and once with *n*-butanol, before the centrifuge-driven dialysis on a centricon 30 (Amicon). In 12 μ l of TE, ~0.5 pmol of PCR product was mixed with 2 pmol of one of the same primers, D6, used in the amplification. This primer had been labelled with ³²P using T4 kinase. The primer-template mixture was heated to 95 °C for 5 min, and immediately placed on ice. Reactions were initiated by adding 2.8 μ l of this mixture to 3.15 μ l of dideoxy G, A, T and C reaction mixtures composed of reagents provided in the 'sequence' kit as follows: 2.5 μ l of the G, A, T or C 'termination mixes', 0.15 μ l of Sequenase (2 units), 0.38 μ l '5 $\times$ buffer' and 0.22 μ l 0.1M DTT. After incubation at 37 °C for 10 min, these reactions were treated as in ref. 13.

Finally, we point out that the DNA content of hairs is usually limited and/or degraded. Nevertheless, as long as there is enough single-stranded DNA to bridge the distance between the 5' ends of the two oligonucleotide primers, PCR amplification can take place (D. Bugawan, unpublished observation). Thus, any biological sample whose DNA content is too low or too degraded for RFLP analysis or molecular cloning, such as small and old blood or sperm stains, histological specimens^{25,26}, or samples of historic or prehistoric interest such as mummified tissue remains^{27,28}, should be amenable to genetic analysis using PCR.

Portions of this research were supported by a grant to G.F.S. from the National Institute of Justice. We thank E. Blake, D. Bugawan, D. Gelfand, B. Griffith, G. Horn, P. Morin, S. Scharf, R. Saiki and Tom White for their help and discussions, D. Gelfand, S. Stoffel, for *Taq* DNA polymerase, D. Spasic, L. Goda, C. Chang and C. Levenson for oligonucleotides, T. Kocher for D-loop amplification primer D18, L. Johnson and the engineering group at Cetus for equipment for PCR, and S. Nilson and E. Ladner for help with figures.

Received 12 January; accepted 1 March 1988.

1. Caskey, C. T. *Science* **236**, 1223-1229 (1987).
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76-79 (1985).
3. Gill, P., Jeffreys, A. J. & Werrett, D. J. *Nature* **318**, 577-579 (1985).
4. Kantner, E. et al. *J. Forensic Sci.* **31**, 403-408 (1986).
5. Giusti, A. et al. *J. Forensic Sci.* **31**, 409-417 (1986).
6. Sensabaugh, G. F. *J. Forensic Sci.* **31**, 393-396 (1986).
7. Wong, Z. et al. *Ann. hum. Genet.* **51**, 269-288 (1987).
8. Saiki, R. K. et al. *Science* **230**, 1350-1354 (1985).
9. Faloona, F. & Mullis, K. *Meth. Enzym.* **155**, 335-350 (1987).
10. Saiki, R. et al. *Science* **239**, 487-491 (1988).
11. Hukkelhoven, M. W. A. C., Vromans, E., Markslag, A. M. G. & Vermorken, A. J. M. *Anticancer Res.* **1**, 341-344 (1981).
12. Cann, R. L. & Wilson, A. C. *Genetics* **104**, 699-711 (1983).
13. Wrischnik, L. et al. *Nucleic Acids Res.* **15**, 529-542 (1987).
14. Kogan, S. C., Doherty, M. & Gitschier, J. *New Engl. J. Med.* **317**, 985-990 (1987).
15. Wallace, R. B. *Nucleic Acids Res.* **6**, 3543-3557 (1979).
16. Saiki, R. K. et al. *Nature* **324**, 163-166 (1986).
17. Wong, C., et al. *Nature* **330**, 384-386 (1987).
18. Aquadro, C. F. & Greenberg *Genetics* **103**, 287-312 (1983).
19. Smith, L. et al. *Nature* **321**, 674-678 (1987).
20. Prober, J. M. et al. *Science* **238**, 336-341 (1987).
21. Bisbing, R. E. in *Forensic Science Handbook* (ed. Saferstein, R.) 185-221 (Prentice-Hall, Englewood Cliffs, NJ 1982).
22. Divall, G. B. *Electrophoresis* **6**, 249-258 (1985).
23. Gerhard, M. & Hermes, M. *Electrophoresis* **8**, 490-492 (1987).
24. Benciolini, P. & Cortino, P. in *Advances in Forensic Science* Vol. 1 (eds Lee, H. C. & Gaensslen, R. E.) 106-114 (Biomedical, Foster City, California, 1985).
25. Ipraim, C. C. et al. *Biochem. biophys. Res. Commun.* **142**, 710-716 (1987).
26. Shibata, D. K., Arnheim, N. & Martin, J. W. *J. exp. Med.* **167**, 225-230 (1988).
27. Pääbo, S. *Nature* **314**, 644-645 (1985).
28. Higuchi, R., Bowman, B., Freiburger, M., Ryder, O. A. & Wilson, A. C. *Nature* **312**, 282-284 (1984).
29. Horn, G. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
30. Sheldon, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9085-9089 (1986).
31. Chien, A., Edgar, D. B. & Trela, J. M. *J. Bacteriol.* **127**, 1550-1557 (1976).
32. Church, G. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
33. Anderson, S. et al. *Nature* **290**, 457-465 (1981).
34. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
35. Thuring, R. W. J., Sanders, J. P. M. & Borst, P. *Analyt. Biochem.* **66**, 213-220 (1975).

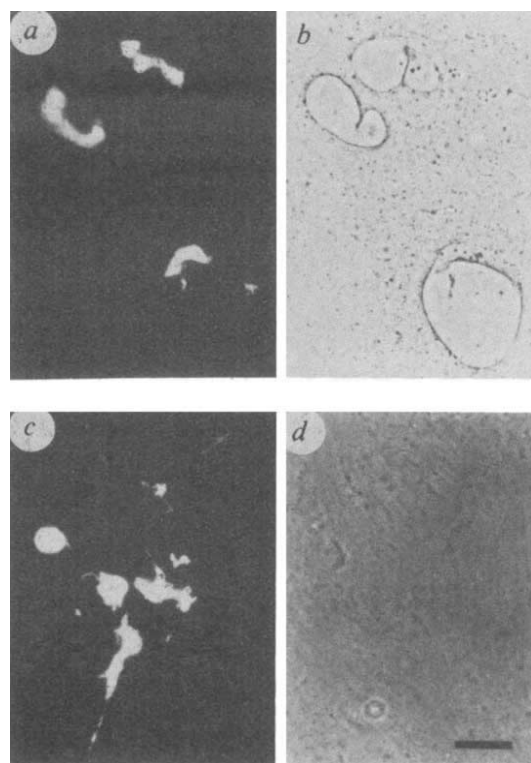


Fig. 1 Protein synthesis permits mitotic changes *in vitro*. Sperm chromatin (3 ng DNA per μ l) was incubated for 5 h in extract plus (a, b) or minus (c, d) 100 μ g ml⁻¹ cycloheximide. Nuclei were examined under Hoechst 33258 UV fluorescence (a, c) or phase contrast optics (b, d). Scale bar 20 μ m; all fields at same magnification.

Methods. Extracts of *Xenopus* eggs were prepared by the method described¹ with the following modifications. Prior to crushing, eggs were packed by centrifugation at 1,500 r.p.m. in an SW50 rotor (Beckman) for 1 min at 4 °C, and all excess buffer was removed. Cytochalasin B (10 μ g ml⁻¹) was added to the extract after crushing. Extract was frozen with 3% v/v glycerol.

A role for the nuclear envelope in controlling DNA replication within the cell cycle

J. Julian Blow & Ronald A. Laskey

CRC Molecular Embryology Research Group,
Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK

In eukaryotes the entire genome is replicated precisely once in each cell cycle. No DNA is re-replicated until passage through mitosis into the next S-phase. We have used a cell-free DNA replication system from *Xenopus* eggs to determine which mitotic changes permit DNA to re-replicate. The system efficiently replicates sperm chromatin, but no DNA is re-replicated in a single incubation¹. This letter shows that nuclei replicated *in vitro* are unable to re-replicate in fresh replication extract until they have passed through mitosis. However, the only mitotic change which is required to permit re-replication is nuclear envelope permeabilization. This suggests a simple model for the control of DNA replication in the cell cycle, whereby an essential replication factor is unable to cross the nuclear envelope but can only gain access to DNA when the nuclear envelope breaks down at mitosis.

When sperm chromatin is incubated in homogenates of *Xenopus* eggs it is assembled into normal interphase nuclei surrounded by a nuclear envelope¹⁻⁴ (Fig. 1a, b). When such

extracts are permitted to synthesize protein (>40 μ g protein per ml extract in 4 h), nuclei assembled *in vitro* subsequently undergo the early mitotic changes of chromatin condensation and nuclear envelope breakdown² (Fig. 1c, d).

In addition, DNA assembled into nuclei *in vitro* is replicated efficiently^{1,5-7}. Bromodeoxyuridine density substitution shows that only a single round of replication occurs in a single incubation¹ in the presence or absence of protein synthesis (Fig. 2a).

The extract distinguishes replicated from unreplicated nuclei. Nuclei replicated in the presence of cycloheximide, and then transferred to fresh extract, do not re-replicate (Fig. 2b, filled symbols and Fig. 3b). If protein synthesis is permitted, however, replicated nuclei slowly acquire the ability to re-replicate in fresh extract (Fig. 2b, open symbols and Fig. 3a). The ability to re-replicate corresponds to the time at which chromatin condensation and nuclear envelope breakdown occur (Figs 1, 2).

To determine if nuclear envelope breakdown and chromatin condensation are the changes that permit nuclei to re-replicate, nuclei were replicated in the presence of cycloheximide and then given various treatments before transfer to fresh extract (Fig. 3c-f). Figure 3c shows the effect of treatment with Maturation Promoting Factor (MPF). This is an activity found in most mitotic or meiotic cells that can induce mitotic changes in interphase nuclei⁹⁻¹¹. Treatment with MPF, which causes nuclear envelope breakdown and chromatin condensation, can substitute for protein synthesis in allowing nuclei to re-replicate when added to fresh extract (Fig. 3a-c).

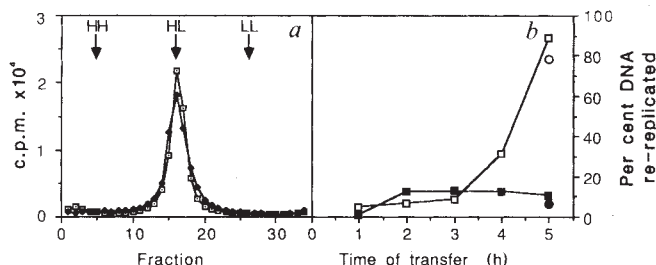


Fig. 2 The extract can distinguish replicated from unreplicated DNA. *a*, Density substitution of sperm chromatin incubated *in vitro* shows a single round of semi-conservative replication with or without cycloheximide. *b*, Time course showing that replicated DNA can be re-replicated on addition to fresh extract only in the presence of protein synthesis.

Methods. *a*, Sperm chromatin (3 ng DNA per μ l) was incubated in egg extract with α -³²P-dATP and 0.4 mM BrdUTP for 5 h; DNA was extracted and fractionated on CsCl density gradients. Density shown by arrows: HH 1.79; HL 1.75; LL 1.71. ($\square$), Control (untreated) extract; ($\blacklozenge$) extract with 100 μ g ml⁻¹ cycloheximide. *b*, Sperm chromatin (3 ng DNA per μ l) was incubated in 20 μ l extract with α -³²P-dATP, 0.5 mM BrdUTP, plus or minus 100 μ g ml⁻¹ cycloheximide, for various times, and then resuspended in 1 ml Buffer A (ref. 5). Nuclei were pelleted at 2,000g for 2 min, and resuspended in fresh extract containing ³H-dATP, 0.5 mM BrdUTP, plus or minus 100 μ g ml⁻¹ cycloheximide. This extract was incubated for 4 h, when DNA was isolated and fractionated on CsCl density gradients¹. ($\square$), First and second extracts untreated (no cycloheximide). ($\blacksquare$), First and second extracts contained 100 μ g ml⁻¹ cycloheximide. Either first ($\bullet$), or second ($\circ$) extracts only contained cycloheximide.

Figure 3d-f shows that nuclear envelope permeabilization alone is sufficient to allow re-replication. Thus when nuclei are replicated in the presence of cycloheximide, and then treated with agents that permeabilize the nuclear envelope, they can re-replicate in fresh extract. Lysolecithin, a lecithin analogue which inserts into lipid membranes and permeabilizes them, permits 43% of the replicated DNA to re-replicate (Fig. 3d). Melittin and phospholipase, which hydrolyses phospholipids in the nuclear envelope, permit 27% of the replicated DNA to re-replicate (Fig. 3e). Similarly, mechanical shear, such as that caused by pelleting the nuclei in a fixed angle rotor, permits replicated DNA to re-replicate (ref. 1 and data not shown). However, degradative treatments that do not permeabilize the nuclear envelope, such as phosphatase (Fig. 3f) and RNase (data not shown) do not allow re-replication. When only a fraction of the template DNA re-replicates (Fig. 3c-e) most nuclei re-replicate either fully or not at all (flow cytometry data, not shown). Therefore nuclear envelope permeabilization during mitosis is sufficient to permit nuclei to re-replicate completely in the subsequent interphase.

We have previously discussed positive and negative regulatory models for limiting DNA replication to a single round per cell cycle^{12,13} but experiments performed during the course of this work exclude aspects of each of them (data not shown). Instead these results can be explained if an essential replication factor is unable to cross the nuclear envelope during interphase and can only gain access to DNA when the nuclear envelope breaks down at mitosis. This model is outlined in Fig. 4. In Fig. 4a this factor ('Licensing Factor') is shown binding to chromatin during mitosis or at the start of an incubation. Before DNA can replicate, it must be assembled into a nucleus with a complete nuclear envelope^{1,4-6,14}. In the model, no unbound Licensing Factor remains free in the nucleoplasm: possibly because it is stable in the nuclear environment only when bound to DNA, or because its only route for inclusion in the nucleus is by binding DNA, nascent nuclear envelope being very closely applied to the surface of the chromatin⁴ (Fig. 4b). Only after the DNA is assembled into a mature nucleus are replication forks initiated

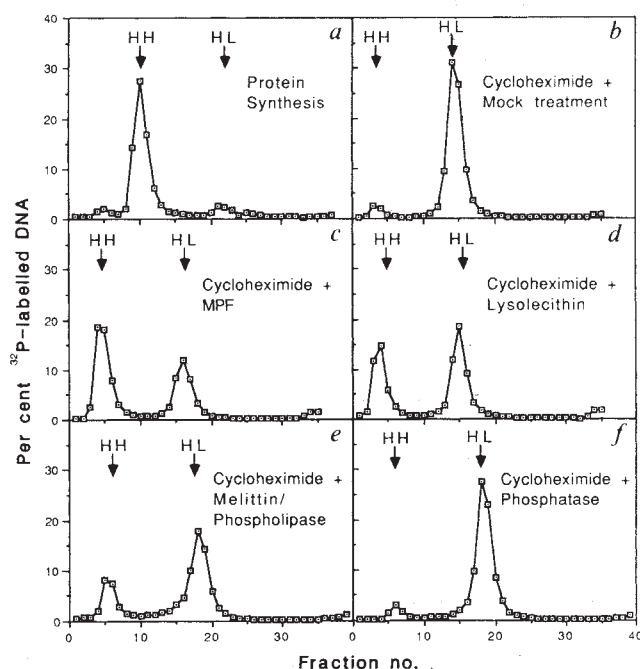


Fig. 3 Nuclear envelope permeabilization between incubation in two successive extracts allows re-replication of replicated nuclei (seen as HH DNA). *a*, Nuclei replicated in the absence of cycloheximide. *b-f*, Nuclei replicated in the presence of cycloheximide and treated with: *b*, mock treatment; *c*, maturation promoting factor; *d*, lysolecithin; *e*, melittin and phospholipase; *f*, phosphatase. Incubation conditions were as for Fig. 2b. The percentage of ³²P-labelled DNA in each fraction is given. The ³H content of each fraction was also measured; this confirmed the identification of the peaks as arrowed.

Methods. Panels *a* and *b*, nuclei were transferred as for Fig. 2b. *c*, Extract was supplemented with 50% volume of MPF¹¹ for 2 h before nuclei were transferred as for Fig. 2b. *d*, Nuclei were pelleted as for Fig. 2b, resuspended in 100 μ g ml⁻¹ lysolecithin (100 μ l) for 10 min, diluted with 400 μ l Buffer A plus 2% BSA, underlayered with fresh extract, and transferred by centrifugation at 5,000 r.p.m. for 2 min in an SW50 rotor (Beckman). *e*, Nuclei pelleted as for Fig. 2b were resuspended in 100 μ l Buffer A plus 50 μ g ml⁻¹ melittin, 50 μ g ml⁻¹ phospholipase A, 1 mM Ca²⁺, 1 mM Mg²⁺ for 10 min, diluted with 400 μ l Buffer A plus 2% BSA, 1 mM EDTA, and transferred as in *d*. *f*, Nuclei pelleted as for Fig. 2b were resuspended in 50 μ l Buffer A pH 8.2 plus 25 units of calf intestine alkaline phosphatase, diluted and transferred as in *d*.

throughout the nucleus⁵ at the sites of bound Licensing Factor (Fig. 4c). Licensing Factor only supports a single initiation event where it is bound to the DNA, and is inactivated by either initiation or the passage of a replication fork (Fig. 4d). This means that the entire genome is replicated precisely once, and no re-replication can occur (Fig. 4e). Nuclei in G2 are unable to re-replicate because Licensing Factor in the cytoplasm cannot gain access to the DNA until the nuclear envelope is permeabilized during mitosis. Thus this model can explain why G2 nuclei in G1/G2 cell hybrids cannot re-replicate until passage through mitosis¹⁵. Alternative models involving the escape of a diffusible inhibitor at mitosis would have to be much more complex than the one presented here.

Neither the *Xenopus* egg nor egg extract require specific DNA sequences for DNA replication^{1,6,13,16,17}, nuclear formation^{1,6,13,18,19} or for limiting DNA replication to a single round per cell cycle^{16,20}. The occasional re-replication of plasmid DNA in incubations *in vitro*¹ probably reflects the fragility of pseudo-nuclei assembled from pure DNA *in vitro*. The *Xenopus* embryo therefore differs from bovine papilloma viruses, which require special *cis*-acting sequences to constrain viral replication to a single round per cell cycle^{21,22}.

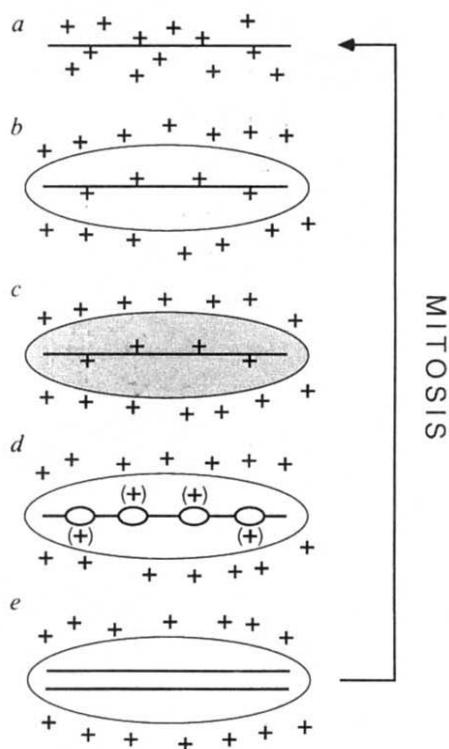


Fig. 4 Model for the control of DNA replication in the *Xenopus* early embryo. *a*, Licensing factor (+) binds to DNA. *b*, DNA is assembled into nucleus. *c*, Initiation of licensed sites occurs coordinately throughout individual nuclei⁵. *d*, Licensing factor is inactivated by initiation or the passage of a replication fork. *e*, Fully replicated DNA cannot re-replicate due to exclusion of licensing factor from DNA by nuclear envelope. Breakdown of the nuclear envelope during mitosis allows access of the Licensing Factor to the DNA, to prepare it for DNA synthesis in the next cell cycle.

Consistent with our model, G2 nuclei from HeLa cells must be prepared using detergent to be capable of replication in the first cell cycle after microinjection into *Xenopus* eggs²³. But many eukaryotic cells, such as yeast, do not undergo nuclear envelope breakdown during mitosis. The model could still operate in these organisms so long as nuclei become permeable to the Licensing Factor during mitosis. The morphological changes of the yeast nucleus during mitosis²⁴ invite the speculation that the model presented here may be applicable to all eukaryotic cells.

We thank David Blow, Steve Dilworth, Colin Dingwall, Richard Harland, Peter Lachman, Moira Sheehan and Jim Watson for their help. This work was funded by the Cancer Research Campaign. J.J.B. is a Beit Memorial Junior Research Fellow.

Received 11 December 1987; accepted 26 February 1988.

1. Blow, J. J. & Laskey, R. A. *Cell* **47**, 577-587 (1986).
2. Lohka, M. J. & Masui, Y. *Science* **220**, 719-721 (1983).
3. Lohka, M. J. & Masui, Y. *J. Cell Biol.* **98**, 1222-1230 (1984).
4. Sheehan, M. A., Mills, A. D., Sleeman, A. M., Laskey, R. A. & Blow, J. J. *J. Cell Biol.* **106**, 1-12 (1988).
5. Blow, J. J. & Watson, J. V. *EMBO J.* **6**, 1997-2002 (1987).
6. Newport, J. *Cell* **48**, 205-217 (1987).
7. Hutchison, C. J., Cox, R., Drepaal, R. S., Gomperts, M. & Ford, C. C. *EMBO J.* **6**, 2003-2010 (1987).
8. Newport, J. & Kirschner, M. W. *Cell* **37**, 731-741 (1984).
9. Miake-Lye, R. & Kirschner, M. *Cell* **41**, 165-175 (1985).
10. Lohka, M. J. & Masui, Y. *Dev. Biol.* **103**, 434-442 (1984).
11. Newport, J. & Spann, T. *Cell* **48**, 219-230 (1987).
12. Laskey, R. A., Harland, R. M., Ernschaw, W. C. & Dingwall, C. in *International Cell Biology 1980-1981* (ed. Schwiager, H. G.) 161-167 (Springer, Berlin, 1981).
13. Blow, J. J., Dilworth, S. M., Dingwall, C., Mills, A. D. & Laskey, R. A. *Phil. Trans. R. Soc. B* **317**, 483-494 (1987).
14. Blow, J. J., Sheehan, M. A., Watson, J. V. & Laskey, R. A. *Cancer Cells* **6**, (Cold Spring Harbor Laboratory, New York, in the press).

15. Rao, P. N. & Johnson, R. T. *Nature* **225**, 159-164 (1970).
16. Harland, R. M. & Laskey, R. A. *Cell* **21**, 761-771 (1980).
17. Mechali, M. & Kearsey, S. *Cell* **38**, 55-64 (1984).
18. Forbes, D. J., Kirschner, M. W. & Newport, J. W. *Cell* **34**, 13-23 (1983).
19. Newmeyer, D. D., Lucocq, J. J., Burglin, T. R. & De Robertis, E. M. *EMBO J.* **5**, 501-510 (1986).
20. Mechali, M., Mechali, F. & Laskey, R. A. *Cell* **35**, 63-69 (1983).
21. Roberts, J. M. & Weintraub, H. *Cell* **46**, 741-752 (1986).
22. Berg, L., Lusky, M., Stenlund, A. & Botchan, M. R. *Cell* **46**, 753-762 (1986).
23. De Roper, A., Smith, J. A., Watt, R. A. & Barry, J. M. *Nature* **265**, 469-470 (1977).
24. Byers, B. in *The Molecular Biology of the Yeast Saccharomyces* (eds Strathern, J. N., Jones, E. W. & Broach, J. R.) 59-96 (Cold Spring Harbor Laboratory, New York, 1981).

The cytoplasmic protein GAP is implicated as the target for regulation by the *ras* gene product

Carmela Calés, John F. Hancock,
Christopher J. Marshall & Alan Hall

Chester Beatty Laboratories, Institute of Cancer Research,
Fulham Road, London SW3 6JB, UK

About 30% of human tumours contain a mutation in one of the three *ras* genes leading to the production of p21^{ras} oncoproteins that are thought to make a major contribution to the transformed phenotype of the tumour^{1,2}. The biochemical mode of action of the *ras* proteins is unknown but as they bind GTP and GDP and have an intrinsic GTPase activity, they may function like regulatory G proteins³⁻⁷ and control cell proliferation by regulating signal transduction pathways at the plasma membrane⁸⁻¹⁰. It is assumed that an external signal is detected by a membrane molecule (or detector) that stimulates the conversion of p21.GDP to p21.GTP which then interacts with a target molecule (or effector) to generate an internal signal. Recently a cytoplasmic protein, GAP, has been identified that interacts with the *ras* proteins, dramatically increasing the GTPase activity of normal p21 but not of the oncoproteins¹¹. We report here that GAP appears to interact with p21^{ras} at a site previously identified as the 'effector' site^{12,13}, strongly implicating GAP as the biological target for regulation by p21.

The biochemical properties of normal and oncogenically activated forms of p21^{ras} have been described using protein derived from *Escherichia coli* expression systems³⁻⁶. We have used a transient expression system to isolate membranes containing high levels of p21^{ras} from COS cells transfected with *ras* complementary DNAs. This has enabled us to study the kinetics of nucleotide exchange and of GTP hydrolysis under more physiological conditions than is possible with protein expressed in *E. coli*. A Western blot of membranes isolated from transfected COS cells (Fig. 1) shows that the characteristic modification of pro-p21 to mature p21 occurs in this system (compare lanes 2 and 5 with 3 and 6) and that ~1% of the membrane protein is p21. Using this approach, it is possible to isolate membranes containing high levels of a variety of mutant *ras* proteins.

There is substantial p21-independent GTP/GDP binding and GTPase activity in the membranes and we have therefore used an immunoprecipitation technique to look at the properties of p21. To determine the GTPase activity of membrane-bound p21, [α -³²P]GTP was preincubated with membranes in the presence of low concentrations of Mg²⁺. Under these conditions exchange on to p21 occurs rapidly (half life, 2 min), but after addition of Mg²⁺, exchange is slowed dramatically (half life, 60 min) and GTP hydrolysis begins⁵. After various time intervals, p21 was solubilized from the membrane, immunoprecipitated, and the ratio of bound GTP to GDP determined by thin-layer chromatography (TLC). Such studies show that the apparent half life for hydrolysis of GTP by membrane-bound normal p21 is 10 min. For the oncogenically activated Val12 protein the half

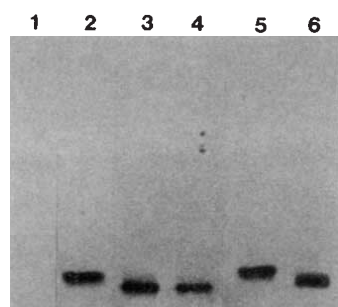


Fig. 1 Western Blot analysis of membranes isolated from COS cells transfected with Ha-ras cDNAs. Tracks are as follows: 1, untransfected COS membranes (20 μ g total protein); 2, *E. coli*-produced p21^{Ha-ras} (200 ng); 3, COS membranes containing p21^{Ha-ras} (20 μ g total protein); 4, COS membranes containing Ala38 p21^{Ha-ras} (20 μ g total protein); 5, *E. coli*-produced Val12 p21^{Ha-ras} (200 ng); 6, COS membranes containing Val12 p21^{Ha-ras} (20 μ g total protein). The membrane-bound forms of p21 move faster than the *E. coli* proteins because of the post-translational addition of palmitic acid to Cys186 (ref. 16).

Methods. The cDNA expression plasmid pcEXV (ref. 17) was used to express Ha-ras cDNA in COS cells. *NdeI/SalI* fragments were isolated from plasmids pSKc-Ha-ras (normal) and pSKT24 (Val12) (ref. 18) and cloned directly into the *SmaI* site of the vector. The effector-site mutations were introduced by oligonucleotide-directed mutagenesis into the normal cDNA cloned in M13mp 19. COS cells were transfected using the DEAE-dextran method¹⁹ with 40 μ g plasmid DNA/4 $\times$ 10⁶ cells, and grown for 3 days. Membranes were prepared by homogenizing the cells in 2.5 ml ice-cold hypotonic buffer (10 mM Tris pH 7.5, 5 mM MgCl₂, 1 mM dithiothreitol, 1 mM PMSF). The post-nuclear supernatant was centrifuged for 10 min at 100,000g and the resulting membrane pellet was washed, resuspended in membrane buffer (50 mM Tris pH 7.5, 50 mM NaCl, 5 mM MgCl₂, 1 mM dithiothreitol, 1 mM PMSF), snap-frozen and stored in liquid nitrogen. For detection of p21 on Western blots, rabbit anti-p21 polyclonal antibody was used, followed by sheep anti-rabbit IgG conjugated to alkaline phosphatase. Bands were visualized after the addition of 5-bromo-4-chloro-3-indoyl phosphate and nitro blue tetrazolium (Promega Inc.). Membrane protein was determined by the Lowry procedure²⁰.

life is 500 min (Fig. 2). The GTPase activity of Val12 p21 corresponds to that previously determined using protein produced in *E. coli*, whereas GTP hydrolysis by membrane-bound normal p21 appears to be five-fold faster than by protein from *E. coli*^{3,4,6} (Fig. 2).

Recently a cytoplasmic protein, GAP, has been identified which increases the GTPase activity of *E. coli*-produced normal p21, but has no effect on Val12 protein¹¹. We have confirmed this observation using membrane-bound p21 (Fig. 2). Addition of a concentrated NIH-3T3 cytoplasmic extract decreases the apparent half life for GTP hydrolysis by membrane-bound normal protein to <2 min under the conditions of the experiment, but has no effect on Val12 p21 (Fig. 2). Washing the membrane preparations in high salt (0.5 M NaCl) increases the half life of the GTPase reaction of normal p21 to around 30 min and we conclude that a major part of the discrepancy between the rates of *E. coli*-produced and membrane-bound normal p21 is accounted for by GAP present in the membrane preparations.

Mutational analysis^{12,13} and comparisons with other G proteins¹⁴ have localized the effector region of ras protein (that is, the region thought to interact with the downstream target molecule) to between amino acids 30 and 40. To examine the consequences of effector site mutations in normal p21 on its interaction with GAP, we initially replaced Asp38 with Ala in the normal H-ras cDNA clone. It has already been shown that such a mutation in the oncogenic Val12 protein inactivates its transforming activity^{12,13}. We first determined the biological

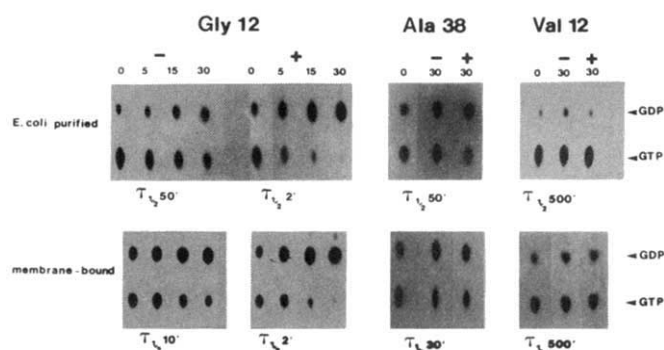


Fig. 2 The effect of GAP on the GTPase activities of normal and oncogenically activated (Val12) p21 and of p21 containing an effector-site mutation (Ala38). Proteins purified from *E. coli* or from COS membranes were loaded with [α -³²P]GTP and the rate of GTP hydrolysis was determined in the absence (–) or presence (+) of cytoplasmic extract containing GAP activity. The times taken for 50% of bound GTP to be converted to bound GDP (half lives) are indicated, though the reaction is only truly first order in the absence of GAP or where GAP has no effect on p21 (that is in Ala38 and Val12 mutants). As GAP appears to act catalytically, the observed half lives for GTP hydrolysis by normal p21 are dependent on the amount of cytoplasmic extract added and represent those measured under the conditions of the experiment, as described below. For Val12 and Ala38, only the 30 min time-points in the presence (+) or absence (–) of GAP are shown.

Methods. Cytoplasmic extract containing GAP activity was prepared from NIH-3T3 cells. Cells were grown to confluence, washed with phosphate-buffered saline (PBS) and detached in PBS containing 0.5 mM EDTA. The cells were lysed in 2 volumes of ice-cold hypotonic buffer and the suspension centrifuged for 30 min at 15,000g. The supernatant (cytoplasmic extract, 5 mg ml^{–1} protein) was dialysed, snap-frozen and stored in liquid nitrogen. To measure the GTPase activity of p21, membranes from transfected COS cells (75 μ g) or p21 produced by *E. coli* (750 ng) were preincubated for 5 min at 30 °C with 1 μ M [α -³²P]GTP (25 Ci mmol^{–1}) in 150 μ l buffer (50 mM HEPES pH 7.8, 150 mM NaCl, 1 mM dithiothreitol, 20 μ M Na₂VO₄, 0.1 mM ATP) with low free Mg²⁺ (10 mM EDTA, 5 mM MgCl₂). The p21 exchange reaction was stopped and the GTPase reaction was initiated by increasing the concentration of free Mg²⁺ to 5 mM. To examine the effect of GAP, cytoplasmic extract (final concentration, 0.5 mg ml^{–1}) was added at *t* = 0. Aliquots (30 μ l) were removed and kept on ice after different time intervals. Membrane-bound p21 was solubilized by adding deoxycholate to a final concentration of 10 mM and immunoprecipitated using the rat monoclonal antibody Y13-259 (ref. 21) and rabbit anti-rat coated proteinA-Sepharose beads. Immunoprecipitates were boiled for 3 min in 10 μ l 0.1% SDS/1 mM EDTA to release bound nucleotide from p21. After centrifugation, the supernatant (5 μ l) was loaded on to TLC plates (PEI-cellulose F, Merck) and chromatographed in 0.75 M potassium phosphate, pH 3.4. GTP and GDP were first visualized by autoradiography and the spots were then cut out of the TLC plate and counted. Because there is a high non-specific phosphatase activity in the membrane preparations, a significant amount of GDP is produced and exchanges on to p21 during the 5 min preincubation period in the low Mg²⁺ conditions. After addition of Mg²⁺, nucleotide exchange is very slow and any increase in the GDP bound to p21 is due to the intrinsic GTPase activity⁵. Background GTP binding was determined by omitting p21-specific antibody from the immunoprecipitations or by using non-transfected COS cells, and amounted to less than 5% of binding in the experiments.

activity of this protein using both transfection of the cDNA into NIH-3T3 and microinjection of the *E. coli*-produced protein into NIH-3T3 cells. As shown in Fig. 3, although wild-type Ha-ras cDNA in the pcEXV-3 vector gave 410 foci per μ g DNA after transfection, Ala38 gave none. (As a comparison, Val12 cDNA gives 20,000 foci per μ g and a Val12/Ala38 double mutant gives <1 focus per μ g; data not shown.) Similarly, microinjec-

	EFFECTOR REGION										GAP ACTIVITY	BIOLOGICAL ACTIVITY (foci per µg)	
	12	30	35	38	40								
wild-type	gly	asp glu tyr	asp pro thr	ile glu asp	ser tyr						+	+	(410)
ala35			ala								-	-	
ala38				ala							-	-	
glu38				glu							-	-	
lys40					lys						-	-	
phe40					phe						+	+	(130)

Fig. 3 Effector site mutations in normal p21^{ras} showing effect on biological activity and interaction with GAP. The effect of GAP activity on membrane-bound proteins was determined as described in Fig. 2 legend. (+), Increase in the GTPase activity of p21 in the presence of GAP; (-), no effect of GAP on GTPase activity. The biological activity of wild-type p21 and effector mutants was determined by transfection of cDNAs in pcEXV-3 vectors into NIH-3T3 cells and scoring for foci after 14 days²².

tion of normal purified p21 into NIH-3T3 cells at 3 mg ml⁻¹ induced maximum DNA synthesis and complete morphological transformation, whereas the purified Ala38 protein was totally inactive, even after injection at 10 mg ml⁻¹ (Val12 p21 gives a maximum response at 0.3 mg ml⁻¹). To confirm that the Ala38 effector mutation had no effect on the intrinsic biochemical properties of the protein, we measured the kinetics of GTP/GDP exchange (data not shown) and the intrinsic GTPase activity (Fig. 2) of the purified *E. coli*-derived protein. These were virtually indistinguishable from normal p21 (refs 4, 5). Addition of GAP activity to the purified Ala38 protein, however, revealed a striking difference from the wild-type protein, there being no detectable increase in its GTPase activity (Fig. 2). This result was confirmed using membrane-bound Ala38 protein (Fig. 2). More recently we have used GAP that was partially purified by FPLC and obtained the same results (data not shown), making it unlikely that components other than GAP in the cytoplasmic extract are required for this effect.

To extend these observations to other amino-acid substitutions in the effector region, we have examined the properties of Thr35→Ala, Asp38→Glu, Tyr40→Lys and Tyr40→Phe. Two of these mutations, Ala35 and Lys40, have previously been reported as destroying the transforming activity of oncogenically activated Val12 p21 (ref. 12). Figure 3 shows that, apart from Phe 40, all these mutations, including the conservative Asp38→Glu38 change, behave like the Ala38 mutation and destroy the transforming activity of normal p21^{H_a-ras}. The GTPase activities of these proteins were determined by transfecting cDNA clones into COS cells. The GTPase half lives of Ala35, Glu38 and Lys40 were all ~30 min, in contrast to the normal protein and Phe40, which were both ~10 min (Fig. 4). Addition of cytoplasmic extract containing GAP activity further reduced the apparent GTPase half life of normal and Phe40 proteins to less than 2 min (Figs 2 and 4), but had no effect on Ala35, Glu38 or Lys40 (Fig. 4). The GTPase half lives of 30 min for these three proteins are essentially the same as that of *E. coli*-produced normal protein measured in the absence of GAP. The GTP/GDP exchange rates for all five effector site mutations described here are identical to that of normal protein (data not shown). Thus, four amino acid substitutions in the effector region which destroy the biological activity of both normal and oncogenically activated p21^{ras} also interfere with the ability of GAP to increase the GTPase activity of the normal proteins. A Phe40 mutation in this region which does not affect the transforming activity does not interfere with GAP.

We conclude that GAP probably interacts with p21^{ras} at the effector site (amino acids 30–40). Possibly GAP interacts at a distinct site that is conformationally affected by mutations in the effector region, though we think this unlikely as the same

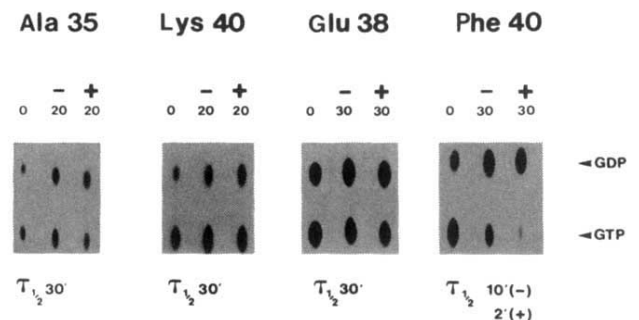


Fig. 4 Effect of GAP activity on p21 effector mutants. Four different effector mutant cDNAs, Ala35, Glu38, Lys40 and Phe40, were transfected into COS cells and membranes isolated as described in Fig. 1 legend. GTPase activities of membrane-bound p21 were determined in the presence (+), or absence (-) of cytoplasmic extract (final concentration, 0.5 mg ml⁻¹), as described in Fig. 2 legend.

mutations have no effect on the intrinsic nucleotide exchange or GTPase activities of p21. Alternatively, GAP may bind at a distinct site and exert its effect on the GTPase activity by inducing conformational changes through the effector region. We cannot exclude this possibility, except to say that an internal deletion towards the carboxy terminus (amino acids 171–184) has no effect on the GAP-stimulated GTPase activity. On the basis of the mutational analysis presented here, we propose that GAP binds at the effector site of p21^{ras} and that it is the target molecule for regulation by p21^{ras}. The association of GAP with normal p21.GTP leads to a conformational change in *ras* protein increasing its intrinsic GTPase activity in a way analogous to the elongation factor EF.Tu whose GTPase activity is markedly stimulated after binding to the target *E. coli* ribosome¹⁵. We predict that GAP must also interact with Val12 and other oncogenic forms of p21, although in these cases the activating mutation in *ras* protein attenuates the induced conformational change and there is no increase in GTPase activity. We believe that this attenuation of a GAP-induced conformational change is the basis of the oncogenic activation of *ras*. GAP is predominantly cytoplasmic, and its interaction with activated p21^{ras} probably involves relocation to the plasma membrane. Indeed, this may be part of the activation process. There is some indirect evidence that p21^{ras} might regulate a phospholipase C activity^{9,10}. However, it will be necessary to purify GAP before any associated enzymatic activities can be assayed. Definitive proof that GAP is the effector molecule for p21^{ras} will have to await cloning and mutational analysis of the gene.

We thank F. McCormick for providing information concerning GAP before publication, and Hugh Paterson and Geraldine Mbamala for performing the DNA transfection and the protein microinjection experiments. This work was supported by the Cancer Research Campaign (UK) and the Medical Research Council (UK). C.C. is the recipient of a long-term EMBO fellowship.

Received 15 January; accepted 1 March 1988.

- Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1986).
- Paterson, H. *et al. Cell* **51**, 803–812 (1987).
- McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. *Nature* **319**, 644–649 (1984).
- Trahey, M. *et al. Molec. cell. Biol.* **7**, 541–544 (1987).
- Hall, A. & Self, A. *J. biol. Chem.* **261**, 10963–10965 (1986).
- Sweet, R. W. *et al. Nature* **311**, 273–275 (1984).
- Hurley, J. B., Simon, M. I., Teplow, D. B., Robishaw, J. D. & Gilman, A. G. *Science* **226**, 860–862 (1984).
- Mulcahy, L. S., Smith, M. P. & Stacey, D. W. *Nature* **313**, 241–243 (1985).
- Wakelam, M. J. O. *et al. Nature* **323**, 173–176 (1986).
- Lacal, J. C. *et al. Science* **238**, 533–535 (1987).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- Sigal, I. S., Gibbs, J. B., D'Alonzo, J. S. & Scolnick, E. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4725–4729 (1986).
- Willumsen, B. M. *et al. Molec. cell. Biol.* **6**, 2646–2654 (1986).

14. McCormick, F. *et al. Science* **280**, 78–82 (1985).
15. Kaziro, Y. *Biochim. biophys. Acta* **505**, 95–127 (1978).
16. Buss, J. E. & Sefton, B. M. *Molec. cell. Biol.* **6**, 116–122 (1986).
17. Muller, J. & Germain, R. J. *exp. Med.* **164**, 1478–1489 (1986).
18. Gross, M. *et al. Molec. cell. Biol.* **5**, 1015–1024 (1985).
19. Lowe, D. G. & Goeddel, D. V. *Molec. cell. Biol.* **7**, 2845–2856 (1987).
20. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1950).
21. Furth, M. E., Davis, L. J., Fleurdelys, B. & Scolnick, E. M. *J. Virol.* **43**, 294–304 (1982).
22. Marshall, C. J., Hall, A. & Weiss, R. A. *Nature* **299**, 171–173 (1982).

Transcriptional but not translational regulation of HIV-1 by the *tat* gene product

Andrew P. Rice & Michael B. Mathews

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

Human immunodeficiency virus-1 (HIV-1), which causes AIDS (acquired immune deficiency syndrome)^{1,2}, possesses an essential gene, *tat*, whose product, acting through the long terminal repeat (LTR) sequences of HIV-1, activates viral genes and replication^{3,4}. The mechanism by which *tat* *trans*-activates HIV genes is unclear. Some studies have reported that an increase in messenger RNA accumulation directed by the HIV-1 LTR can explain the action of *tat*^{5,6}, but others suggest that this increase in mRNA levels can only partially explain *trans*-activation, and that translational control mechanisms may also be involved^{7–10}. To test those possibilities we have established an efficient adenovirus system for delivering the HIV-1 LTR attached to a reporter gene (chloramphenicol acetyltransferase; CAT) into cells and monitoring its activity. The HIV-1 LTR expressed from this adenovirus responds to *trans*-activation in a HeLa cell line constitutively expressing the *tat* protein⁷ by increasing the transcription rate of the HIV-1 LTR and the accumulation of mRNA encoding CAT. In this system the translational efficiency of this CAT mRNA in the cell is unaffected by the presence of *tat*.

The use of an adenovirus vector to study *trans*-activation of the HIV-1 LTR by *tat* offers several advantages over the plasmid transfection techniques used previously. An adenovirus vector allows the highly efficient introduction of the HIV-1 LTR into a variety of cell types of rodent, simian, and human origin; it also allows experiments to be scaled to preparative levels for biochemical analysis, and gives access to the extensive repertoire of regulatory mechanisms found in adenovirus. For our present purposes, we constructed two recombinant adenoviruses. To make the first, the HIV-1 LTR, fused to *CAT*, was removed from a bacterial plasmid¹¹ and incorporated into adenovirus by *in vivo* recombination¹². This virus, termed HIV-1CAT-ad, contains nucleotides –453 to +80 (cap site +1) from the 3' HIV-1 LTR. These sequences include all of the HIV-1 LTR regulatory elements which have been defined, including a *tat* responsive element known as TAR (–17 to +54)^{6,13}. The TAR element contains sequences transcribed into mRNA, suggesting that the *tat* protein may interact with RNA. As a control, we also constructed a second, similar adenovirus, RSVCAT-ad, containing the Rous sarcoma virus (RSV) LTR fused to *CAT*. The RSV LTR is insensitive to *tat*^{9,13}. These two viruses lack most of the adenovirus *E1* sequences and do not produce any of the multiple products of the *E1A* and *E1B* genes¹⁴, including the 13S *E1A* gene product whose expression is required for *trans*-activation of all other early adenovirus promoters^{15,16}. In consequence, transcription of adenovirus genes is extremely inefficient in these viruses. Virus stocks are therefore propagated and titered in human 293 cells, which are transformed by the left-hand end of adenovirus and supply the *E1* gene products *in trans*¹⁷,

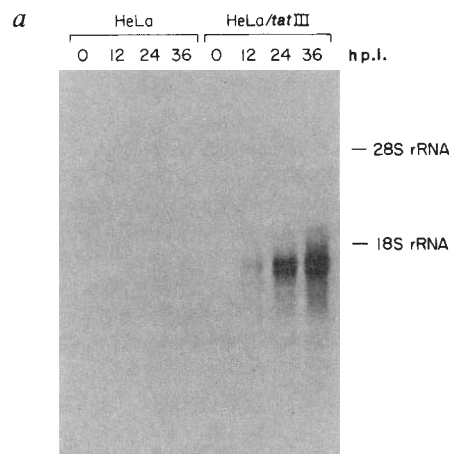


Fig. 1 CAT mRNA accumulation after infection. Total cytoplasmic RNA was extracted from HeLa and HeLa/*tat* cells after infection with HIV-1CAT-ad and CAT mRNA detected by Northern blot analysis¹⁹ using a CAT anti-sense RNA ³²P-probe. CAT mRNA extracted from HIV-1CAT-ad-infected cells migrates slightly faster than 18S ribosomal RNA (1,900 nucleotides (nt)), in agreement with the predicted size (1,700 nt). RNA was extracted at the indicated time points after infection with HIV-1CAT-ad at multiplicity of infection (MOI) of 300; 10 µg cytoplasmic RNA were loaded in each lane of a 1.0% formaldehyde/agarose gel. Migration of 28S and 18S rRNA was visualized by UV shadowing of the gel before transfer. Shown is a 15-h exposure of the filter to X-ray film using an intensifying screen at –70 °C.

allowing easy preparation of high-titre virus stocks.

To determine if the HIV-1 LTR located within an adenovirus vector responds to *trans*-activation by *tat*, we used a HeLa cell line that constitutively expresses the *tat* protein. This cell line, called HeLa/*tat*, was made by co-transfection of a neomycin-resistance plasmid and a *tat*-expressing subclone of HIV-1 provirus⁷. We have found that expression of CAT activity following RSVCAT-ad infection of HeLa/*tat* cells and the parental HeLa cells (both obtained from W. Haseltine and colleagues, Harvard Medical School) is equivalent in the two cell lines¹⁸. Only a small amount of CAT activity is detected following HIV-1CAT-ad infection of HeLa cells; however, several-hundredfold higher levels are observed following HIV-1CAT-ad infection in HeLa/*tat* cells. Thus the HIV-1 LTR responds specifically to *trans*-activation by *tat* when inserted into the adenovirus genome.

To examine the mechanisms involved, we first analysed steady-state levels of CAT mRNA following HIV-1CAT-ad infection of HeLa and HeLa/*tat* cells (Table 1). Total cytoplasmic RNA was extracted at various times after infection, and CAT mRNA sequences were detected by Northern blot analysis using a CAT anti-sense RNA probe (Fig. 1). In HeLa cells, CAT mRNA is just barely detectable at 36 h post infection (p.i.). In HeLa/*tat* cells, there is a clear increase in CAT mRNA with time after infection. The large difference in CAT mRNA levels between the two cell lines makes relative quantitation difficult, but we estimate that at 36 h p.i., HeLa/*tat* cells contain over 100-fold more CAT mRNA than HeLa cells.

The very low level of CAT mRNA produced in the absence of *tat* might hamper the examination of possible effects of *tat* on translation of mRNA transcribed from the HIV-1 LTR, but CAT expression can be increased by co-infection of HIV-1CAT-ad or RSVCAT-ad with wild-type adenovirus¹⁸. This effect requires the 13S *E1A* gene product which is known to *trans*-activate a variety of genes at the transcriptional level²⁰. CAT mRNA levels in such co-infections were analysed by Northern blot hybridization and quantified by dot-blot hybridization (Fig.

2a). The infection was carried out with cytosine arabinoside added to the culture medium to block viral DNA replication and avoid complications due to increased template number and the onset of the late phase of infection. In HeLa cells infected with HIV-1CAT-ad alone, no CAT mRNA was detected, in agreement with Fig. 1. CAT mRNA was seen in HeLa cells coinfecting with wild-type adenovirus (ad2) and HIV-1CAT-ad. In HeLa/tat cells, co-infection increased the CAT mRNA level sixfold over that seen in HIV-1CAT-ad infection alone. In both cell lines, co-infection of RSVCAT-ad and wild-type adenovirus increased CAT mRNA level eightfold over those seen in RSVCAT-ad infection alone. Thus, co-infection with adenovirus leads to an increased basal level of CAT mRNA which is not promoter specific, but which in the HIV-1 case can be further augmented by the action of the tat protein.

We exploited this observation to test the hypothesis that tat is additionally involved in translational control⁷⁻¹⁰. To measure translational efficiency directly, we compared the rate of synthesis of CAT protein to the level of CAT mRNA. Parallel cultures to those of Fig. 2 were pulse-labelled for 1 hour with [³⁵S]methionine and synthesis of CAT protein was measured by quantitative immunoprecipitation (Fig. 2b). The results were used to calculate the ratio of CAT protein synthesis to CAT mRNA level (Table 1). As noted above, it is difficult to draw

Table 1 CAT activity after infection

Infection	Relative CAT mRNA levels	
	HeLa	HeLa/tat
HIV-1CAT-ad	ND	1
HIV-1CAT-ad+ad2	2	6
RSVCAT-ad	4	3
RSVCAT-ad+ad2	32	24
Infection	Relative level of synthesis of CAT protein	
	HeLa	HeLa/tat
HIV-1CAT-ad	ND	1.5
HIV-1CAT-ad+ad2	1	3
RSVCAT-ad	4	2
RSVCAT-ad+ad2	8	5
Infection	Ratio of synthesis of CAT protein/mRNA level	
	HeLa	HeLa/tat
HIV-1CAT	ND	1.5
HIV-1CAT-ad+ad2	0.5	0.5
RSVCAT-ad	1	0.7
RSVCAT-ad+ad2	0.25	0.2

The CAT mRNA level in Fig. 2a in HIV-1CAT-ad-infected HeLa/tat cells was arbitrarily defined as 1 unit. For the other RNA preparations, the relative CAT mRNA levels were estimated by choosing the serial dilution which gave a hybridization signal equivalent to this unit. For the rates of synthesis of CAT protein, the immunoprecipitation signal in Fig. 2c from HeLa cells co-infected with HIV-1CAT-ad and wild-type ad2 was arbitrarily defined as 1 unit. To quantify immunoprecipitations, the serial dilution was estimated which gave an equivalent signal to this unit. Immunoprecipitates for HeLa and HeLa/tat cells infected with HIV-1CAT-ad alone were analysed on a separate gel. ND, not determined.

conclusions about effects of tat on translational control from infections with HIV-1CAT-ad alone, but in co-infections of HIV-1CAT-ad and wild-type adenovirus, the translational efficiencies are equivalent in both cell lines. We conclude, therefore, that the translation of CAT mRNA containing TAR sequences is not affected significantly by the presence of tat. The same applies to expression from the RSVCAT-ad virus, whether measured in the presence or absence of co-infecting wild-type adenovirus. (Interestingly, co-infection with wild-type virus seemed to reduce translational efficiency ~3-fold in all cases measured. The reasons for this are unclear, but clearly do not correlate with the presence of tat.) Thus, under all conditions of infection, the efficiency was unaffected by tat.

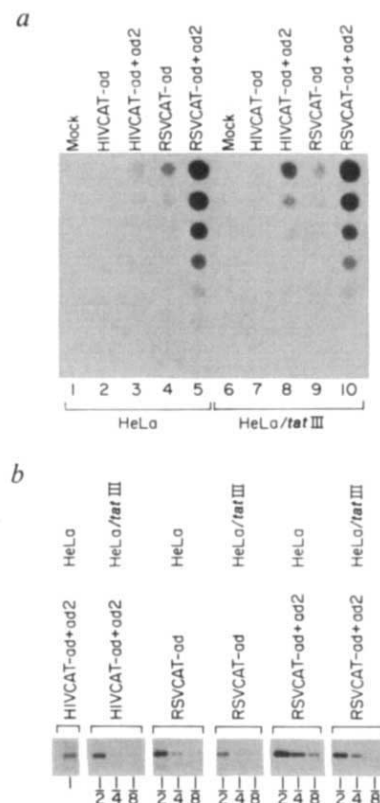


Fig. 2 Comparison of rate of synthesis of CAT protein with CAT mRNA level during infections in HeLa and HeLa/tat cells. HeLa and HeLa/tat cells were infected as indicated. Cytosine arabinoside was added at $20 \mu\text{g ml}^{-1}$, final concentration at 4 h p.i., and 12 h later fresh medium containing the drug was added. At 24 h p.i. culture dishes were labelled for 1 h with [³⁵S]methionine, and synthesis of CAT protein determined by immunoprecipitation; during the labelling period, parallel culture dishes were used to isolate total cytoplasmic RNA. *a*, Dot-blot analysis of CAT mRNA. The RNA preparations were assayed by dot blot hybridization using an anti-sense CAT ³²P-RNA probe. Twofold serial dilutions of RNA were applied to the nitrocellulose filter, beginning with $10 \mu\text{g}$ cytoplasmic RNA. Shown is a 4-h exposure of the filter to X-ray film using an intensifying screen at -70°C . *b*, Rate of synthesis of CAT protein. Cells were labelled for 1 h with [³⁵S]methionine at $300 \mu\text{Ci ml}^{-1}$, and cell extracts were prepared and immunoprecipitated with a polyclonal antiserum to CAT (5 Prime-3 Prime, Inc., Paoli, Pennsylvania). Immunoprecipitates were analysed in 10% SDS-polyacrylamide gels. To assure quantitative immunoprecipitations, two cycles of immunoprecipitation were carried out for each cell extract, and negligible immunoprecipitate products were seen in the second cycle (not shown). For quantitation, twofold serial dilutions of products from the first cycle of immunoprecipitation were analysed. Shown is the region of the gel where CAT protein migrated. The amount of CAT protein immunoprecipitated from $\sim 3 \times 10^5$ HeLa cells co-infected with HIV-1CAT-ad and wild-type ad2 is represented as 1. For other infections, twofold serial dilutions (denoted as 1/2, 1/4, 1/8) of CAT protein immunoprecipitated from 3×10^5 cells were analysed on the gel. Shown is a three-day exposure at -70°C of a fluorographed gel.

This finding implies that the elevated CAT expression from the HIV-1 LTR in the presence of tat is due to regulation of mRNA abundance rather than to translational control. To determine if the increased in mRNA level caused by tat is associated with regulation of transcription from the HIV-1 LTR, we isolated nuclei from cells at 24 h p.i. and performed nuclear run-off assays (Fig. 3). No detectable transcription of the HIV-1 LTR promoted CAT gene is seen in HeLa cells infected with HIV-

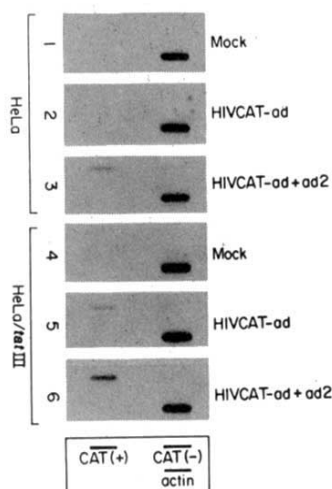


Fig. 3 Transcription rate of HIV-1 LTR in nuclei isolated after infection in HeLa and HeLa/*tat* cells. Cells were infected with HIV-1CAT-ad (MOI 300) or HIV-1CAT-ad plus wild-type adenovirus 2 (HIV-1CAT-ad + ad2, MOI of 50 for each virus) in the presence of cytosine arabinoside as described in Fig. 2b. Nuclei were isolated at 24 p.i. and run-off assays performed and hybridized exactly as described in Greenberg and Ziff²¹. Run-off reactions were hybridized, as diagrammed, to a plasmid clone of β -actin²², and single-strand M13 clones containing CAT sequences that will detect sense (+) CAT RNA or anti-sense (-) CAT RNA. Actin sequences in the plasmid vector are 2,000 nt, CAT sequences in the M13 vectors are 780 nt. Shown is a three-day exposure of the filter to X-ray film with an intensifying screen at -70°C .

1CAT-ad alone (panel 2). As expected, co-infection with wild-type adenovirus activates transcription of the HIV-1 LTR promoted CAT gene (panel 3). A small amount of anti-sense CAT transcription is seen in this co-infection in HeLa cells. In HeLa/*tat* cells, HIV-1 LTR promoted CAT transcription is seen in infections with HIV-1CAT-ad alone (panel 5), demonstrating that *tat* is involved in regulating the transcription of the HIV-1 LTR. In HeLa/*tat* cells co-infected with wild-type adenovirus, the HIV-1 LTR is 'super'-trans-activated at the transcriptional level (compare panels 5 and 6). As before, cytosine arabinoside was used to prevent complications from viral DNA replication: similar results were obtained in the absence of cytosine arabinoside, but in co-infection with wild-type adenovirus we observed high levels of anti-sense CAT transcription, presumably due to the activity of a viral promoter 5' to these sequences. These results indicate that an adenovirus gene product and *tat* are able to act in an additive manner in regulating the level of transcription. The magnitude of the increase in transcription rate brought about by *tat* is comparable to the increase in CAT mRNA level, at least in the co-infection with wild-type adenovirus where basal levels can be measured. In this situation, then, there is little reason to invoke *tat* effects on post-transcriptional mechanisms. It remains to be seen whether the same is true in other systems, where, however, it may not be feasible to make accurate measurements at the different levels of gene expression.

Using an adenovirus vector system, we have found no evidence for translational control by *tat*. Our conclusion disagrees with those based on experiments using transfection of plasmids, suggesting that *tat* regulates translation⁷⁻¹⁰. Unlike the experiments presented here (Fig. 2), these studies generally measured accumulation of protein, rather than its rate of synthesis. Measurements of protein accumulation can only provide an indirect estimate of translational efficiency, and would be misleading if the stability of the protein is greater than that of its mRNA, or if mRNA abundance affects stability. On the other hand, we have found positive evidence for *tat* regulation of transcription of the HIV-1 LTR. A similar result has been

reported in experiments using co-transfections of plasmids in COS cells²³. In that system, however, *tat* was shown to overcome a termination of transcription rather than to regulate initiation of transcription²⁴.

We thank W. Haseltine and colleagues for cell lines and plasmids; S. Rice, L. Shan Su and D. Knipe for CAT M13 clones; E. White for help in construction of recombinant adenoviruses; A. Kamenick for technical help; B. Stillman and W. Herr for critical readings of the manuscript. The work was supported by the National Cancer Institute.

Received 22 October 1987; accepted 23 February 1988.

- Barre-Sinoussi, F. *et al. Science* **220**, 868-871 (1983).
- Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
- Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* **44**, 941-947 (1986).
- Fisher, A. G. *et al. Nature* **320**, 367-371 (1986).
- Peterlin, B. M., Luciw, P. A., Barr, P. J. & Walker, M. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9734-9738 (1986).
- Muesing, M. A., Smith, D. H. & Capon, D. J. *Cell* **48**, 691-701 (1987).
- Rosen, C. A. *et al. Nature* **319**, 555-559 (1986).
- Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807-817 (1986).
- Cullen, B. R. *Cell* **46**, 973-982 (1986).
- Wright, C. M., Felber, B. K., Paskalis, H. & Pavlakis, G. N. *Science* **234**, 988-992 (1986).
- Sodroski, J. *et al. Science* **227**, 171-173 (1985).
- Gluzman, Y., Reich, H. & Solnick, D. in *Eukaryotic Viral Vectors* (ed. Gluzman, Y.) 187-192 (Cold Spring Harbor, New York, 1982).
- Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* **41**, 813-823 (1985).
- DNA Tumor Viruses* (ed. Tooze, J.) (Cold Spring Harbor, New York, 1980).
- Montell, C., Fisher, E. F., Caruthers, M. H. & Berk, A. J. *Nature* **295**, 380-384 (1982).
- Moran, E. & Mathews, M. B. *Cell* **48**, 177-178 (1987).
- Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. J. *gen. Virol.* **36**, 59-72 (1977).
- Rice, A. P. & Mathews, M. B. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5206 (1980).
- Berk, A. J. *Rev. Genet.* **20**, 45-79 (1986).
- Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
- Gunning, P. *et al. Molec. cell. Biol.* **3**, 787-795 (1983).
- Hauber, J., Perkins, A., Heimer, E. P. & Cullen, B. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6364-6368 (1987).
- Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. *Nature* **330**, 489-493 (1987).

Adenovirus E1A products suppress myogenic differentiation and inhibit transcription from muscle-specific promoters

Keith A. Webster, George E. O. Muscat & Larry Kedes

The Medigen Project, Stanford University School of Medicine and Veterans Administration Medical Center, 151M, 3801 Miranda Avenue, Palo Alto, California 94304, USA

The primary function of the adenovirus E1A-region genes is to activate other adenoviral genes during a permissive viral infection by modifying the host cell transcriptional apparatus^{1,2}. Host cell immortalization, or transformation by the whole adenoviral early region, presumably results as a consequence of these modifications. Both transcriptional activation and transcriptional repression of non-adenoviral genes by the E1A proteins have been reported³⁻⁹. It is currently not clear which, if either, of these activities contributes to host cell transformation and immortalization. Although there may be a physiological impact of some E1A-stimulated host cell genes¹⁰⁻¹², in many cases the functional significance is unclear. No common target sequences have been recognized in stimulated cellular genes and it has recently been proposed that in many cases, particularly involving newly transfected genes, available 'TATA-box' sequences may be the opportunistic beneficiaries of E1A assistance as a secondary consequence of E1A primary functions within the host cell nucleus^{1,13}. E1A-mediated transcriptional repression appears to be a more specific process insofar as common core elements are shared by the E1A-suppressed SV40, polyoma B, IgG heavy-chain and insulin enhancers⁹. In the present com-

munication we report that the complete myogenic programme of L8 and C2 myoblasts can be blocked by the introduction of constitutively expressing E1A genes, and show that the transcriptional induction of muscle-specific genes is inhibited. In particular, the promoter-inducing activities of well-defined elements that are required for the muscle-specific expression of the two sarcomeric α -actins, and which normally bind cellular *trans*-acting factors, become targets for E1A suppression. The results support the hypothesis that the suppression of differentiation by E1A products is effected by an E1A-mediated block in the transcriptional activation of cellular genes by specific developmentally regulated *cis*-acting promoter elements.

Rodent L8 and C2 myoblasts were transfected with fragments of adenovirus 5 DNA containing whole early region (E1A + E1B), E1A alone or fragments expressing Ad (adenovirus) 12S or 13S messenger RNAs (generous gifts from E. Ziff¹⁴), along with drug-resistance markers. Neomycin or aminopterin-resistant foci which had been cotransformed with E1A fragments could be immediately recognized by their inability to fuse, a characteristic early step in myoblast differentiation. Non-fusing and fusing foci were counted to estimate the transformation efficiencies of the individual Ad DNA fragments. The individual 12S and 13S fragments transformed myoblasts with approximately equal efficiencies, tenfold lower than that of the whole E1A fragment (as estimated by the ratios of non-fusing/fusing foci per μ g DNA (data not shown)). Five individual clones were selected from each transfection, expanded to mass culture and subjected to conditions that would normally induce fusion and differentiation of myoblasts (see Fig. 2 legend). Figure 1 shows the morphologies of cultures from the indicated L8 and C2 transfections after this treatment. The clones shown displayed the most common transformed morphology. Some of the original clone-isolates unexpectedly displayed the fusion⁺ phenotype in mass cultures. We have not yet determined whether this is dose related or due to unstable integration of the E1A genes. Consequently, transformed lines were subjected to at least 20 passages after removal from selection to ensure phenotypic stability before biochemical analysis.

Further characterization of the stable transfectants in this study was restricted to the L8 lines and data is shown only for the clones transfected with whole early region or complete E1A fragment. Clones expressing only 12S or 13S E1A mRNAs responded identically to those expressing the complete E1A fragment. Figure 2 shows RNA blot analyses of control and transfected L8 lines. Figure 2a and b shows that the clones were expressing Ad E1A or E1A + E1B mRNAs appropriately and that E1A mRNA expression remained essentially constant during exposure of cells to differentiation inducing conditions. Figure 2 (c-e) show the patterns of expression of muscle-specific and non-muscle-specific mRNAs when each of the cell lines was subjected to conditions inductive to myoblast differentiation. Both Ad-DNA transformed lines were dramatically restricted in their ability to induce any of the three contractile protein mRNAs, skeletal α -actin, cardiac α -actin or myosin heavy chain (other muscle-specific transcripts for creatine kinase, carbonic anhydrase III, and troponin T slow isoform were also undetectable in the transformed lines (data not shown)). Mature mRNA at a level approximately 1% of the control was detectable only for cardiac α -actin. No mature skeletal α -actin or myosin heavy chain mRNAs were detectable in the E1A-expressing lines, but in both cases faint bands of lower mobility were apparent upon long exposure of the filters. These RNAs migrated at rates predicted for their respective precursor transcripts. The low level appearance of the mature cardiac α -actin and putative precursor skeletal α -actin and myosin heavy chain mRNAs suggests that the developmental activation of these genes is not completely blocked. The patterns of expression of the non-muscle-specific genes, shown in Fig. 2 (f and g) are qualitatively and quantitatively similar in each line. Typically in L8 wild-type cells the cytoplasmic γ - and β -actin mRNAs are expressed at high levels

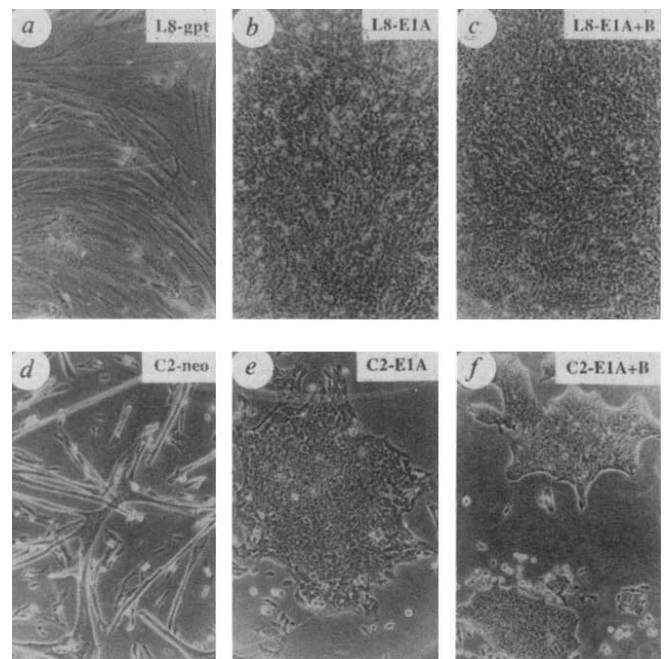


Fig. 1 Inhibition of myoblast fusion by Ad-early genes. (a-c), L8 lines, stably transfected with the indicated genes (pSVE1A¹⁴ and pLB206³⁰) were removed from selection (passage number ~20), grown to confluence and subjected to differentiation inducing conditions (2% horse serum, Dulbecco's modified Eagle medium for six days). Myotubes are clear in the guanosine phosphoribosyl transferase (gpt) transfected culture but not in the adenoviral-DNA transfected lines. (d-f), C2 cells following transfection with the indicated DNAs, panels show drug resistant foci of newly selected cells, subjected to differentiation media for three days. Myotube formation was clearly inhibited in the adenoviral DNA transformed cultures. Neo, neomycin resistant; gpt, aminopterin resistant.

before fusion and gradually decline to 20–30% of their original levels as the cells differentiate. The E1A-expressing lines also show a moderate decline in the levels of these transcripts, although their initial levels are somewhat reduced. In similar assays the expression of the housekeeping genes pyruvate kinase and α -tubulin were minimally effected by adenoviral DNA transformation (data not shown).

Fusion is not a necessary prerequisite for biochemical differentiation, but exit from the cell cycle always precedes activation and expression of muscle specific genes (refs 15–17 and refs therein). Analyses of rates of cell proliferation and the incorporation of [³H]thymidine into DNA while cells were under differentiation-inducing conditions demonstrated that the transformed lines behaved similarly to the wild-type cells (data not shown). The main difference was that the transformed cells became contact-inhibited and ceased [³H]thymidine incorporation at higher cell densities, whereas the wild-type cells fused under these conditions. After eight days in fusion media the non-proliferating Ad DNA-transformed cells formed adhesive sheets and could be maintained in a non-proliferative but viable state for up to two weeks. It therefore appears that these cells exit the cell cycle but are unable to progress to biochemical differentiation.

Previous work from this and other laboratories has shown that activation of the muscle-specific skeletal and cardiac α -actin genes in differentiating myotubes involves the induction or activation of a *trans*-acting transcription factor(s) concomitant with the activation of the respective genomic DNA sequence^{18–20}. In the L8 cells both of these events take place *in vitro* two to three days after the onset of myoblast fusion. In the mouse myogenic line C2, however, the *trans*-acting factor(s) is functionally active

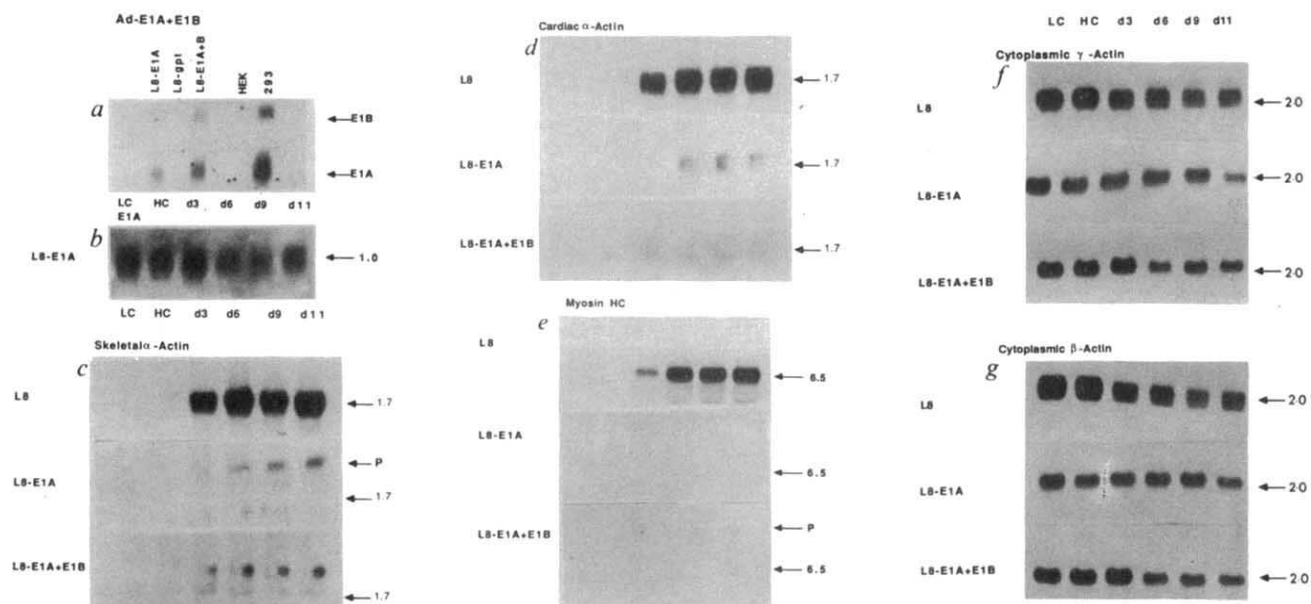


Fig. 2 Expression of muscle-specific and housekeeping mRNAs under differentiation-inducing conditions. *a*, Northern blot of RNAs isolated from confluent cultures of L8 myoblasts stably transfected with the guanosine phosphoribosyl transferase (gpt) selectable marker plus adenovirus 5 E1A or E1A+E1B genes, probed with Ad early-region DNA. Control lanes show human embryonic kidney (HEK) and 293 cells (293 cells constitutively express E1A and E1B mRNAs) to indicate migration of the E1A and E1B mRNAs. *b*, Constitutive expression of E1A mRNA during exposure of confluent L8-E1A cultures for 11 d in low mitogen (2% horse serum, 98% DMEM), differentiation medium. RNAs were probed with 32 P-E1A DNA. LC, low (15%) confluent cultures, HC, high (90%) confluent cultures, d 3–11 are days in differentiation medium. *c*–*g*, Northern blots of RNAs from normal and adenoviral-DNA transfected cells exposed to differentiation medium as above. L8-E1A and L8-E1A+E1B are pools of three Ad-RNA-expressing clones. Complementary DNA probe is indicated at left above each autoradiogram. Arrows indicate transcript size in kilobase pairs, P indicates position of putative precursors.

Methods. RNA extraction and Northern blot procedures, full-length cDNA and nick-translations have been described³¹. Cells, RNA, and hybridization procedures were run in parallel so that the band intensities are quantitatively comparable for each probe.

in proliferating myoblasts before differentiation and gene activation is initiated¹⁹. These myoblasts are therefore fully competent in the tissue-specific transient expression of hybrid gene constructs which contain skeletal²¹ or cardiac^{19,22} α -actin promoters linked to a heterologous reporter gene, in this case the chloramphenicol acetyltransferase gene. These same constructs are not expressed above basal levels in proliferating L8 myoblasts or non-muscle cells¹⁹. This capacity of the C2 cells was used to study directly the interactions of E1A products with the muscle-specific α -actin promoter-constructs by transient expression assays in a situation which is inherently independent of the potential effects of E1A products on cell proliferation. Figure 3 shows the results of these assays in which increasing amounts of pSVE1A (which expresses both E1A genes) were cotransfected with different promoter-CAT expression plasmids into proliferating C2 cells. Expression directed by the skeletal and cardiac α -actin promoters was dramatically inhibited, even at the lower pSVE1A concentrations, whereas that from cytoplasmic actin, or mouse mammary tumour virus (MMTV) promoters was minimally affected. A similar inhibition of the muscle-specific actin promoter activities was observed when these promoter-constructs were cotransfected with pSVE1A into L8 myoblasts and the cultures were induced to differentiate (data not shown). The strong inhibition of muscle-specific, and weak inhibition of cytoplasmic actin promoters parallels the endogenous expression of these genes in the stable E1A-expressing cells (compare with expression of γ - and β -actin in Fig. 2). Cotransfection of pSVE1A and the promoter-constructs into HeLa cells also resulted in a slight inhibition of the γ - and β -actin promoters, comparable with that seen in the C2 cells. The α -actin promoters, which do not express above basal level in HeLa cells, were slightly induced when cotransfected with pSVE1A into these cells (data now shown).

We have previously reported the regions within the skeletal and cardiac α -actin promoters that are important for promoter

activity^{18,19,21,22}. Both promoters contain decameric consensus core sequences CC(A/T)₆GG, (CARg box)¹⁸ which are critical for high level expression from these promoters in muscle cells. These core elements have been shown by *in vitro* analysis to interact with cellular proteins (T. Gustafson, T. Miwa, L. M. Boxer and L. Kedes, manuscript submitted). Figure 4a and b shows schematic maps of the skeletal and cardiac α -actin promoters and indicates the fragments that were linked to the CAT gene and cotransfected with pSVE1A into C2 myoblasts. The transient expression of the CAT gene under the direction of each of these promoter fragments after transfection alone, or with pSVE1A in the ratios 1:5 and 1:1, is shown below each map. The amount of extract tested and time of incubation was varied to enable clear demonstration of the impact of E1A expression (see Fig. 4 legend). The skeletal α -actin promoter contains four CARg box elements, but only the proximal box (most 3'), appears to be important for activity in C2 cells²². Skeletal α -actin promoter-constructs 1–5 (Fig. 4a) all contain this proximal CARg box and all are repressed by pSVE1A cotransfection. This includes the smallest fragment, construct 5, which contains only 66 base pairs (bp), 10 bp of which constitutes the skeletal α -actin proximal CARg box. Clearly the activity of a site within this fragment is sensitive to E1A repression. Gel retardation and dimethyl sulphate protection assays *in vitro* have demonstrated that an intact CARg element is required for a cellular protein(s) to bind specifically to this fragment (T. Gustafson, T. Miwa, L. M. Boxer and L. Kedes, manuscript submitted).

Our previous studies have shown that the proximal CARg box element of the skeletal α -actin promoter increases expression of this promoter tenfold above basal expression in C2 cells. In addition, at least one upstream region acts synergistically with the proximal region to increase expression a further tenfold²². The dramatic repression of constructs 2 and 3, suggests that E1A coexpression eliminated the activity of this additional

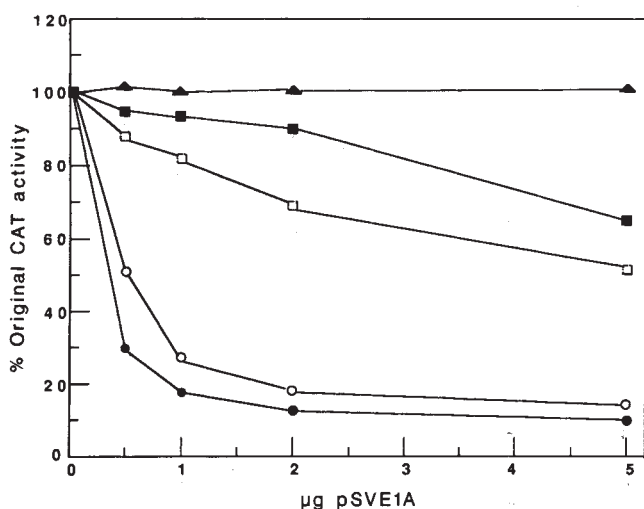
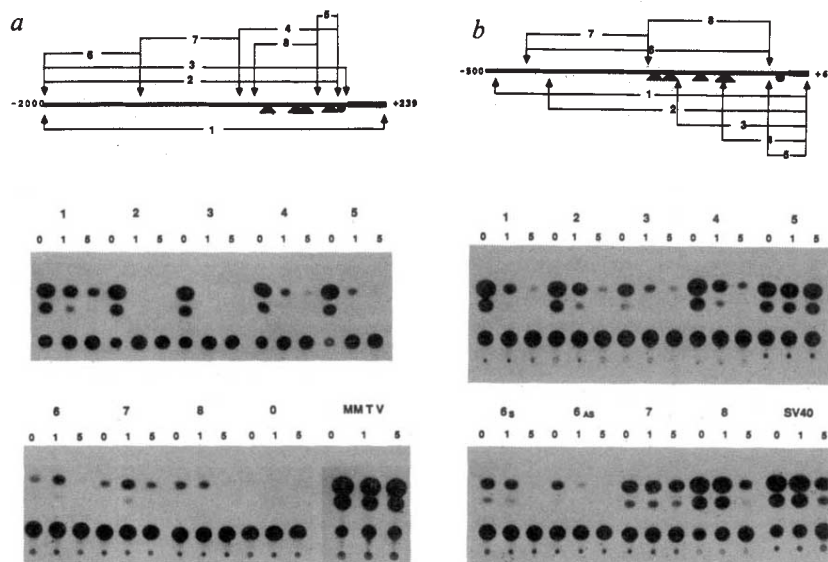


Fig. 3 Transient expression from muscle-specific and non-muscle-specific promoters with cotransfected pSVE1A. Plasmid constructs containing γ -actin, β -actin, α -cardiac actin and α -skeletal actin promoters fused to the CAT gene have been described^{17,18}. A constant amount (5 μ g), of the test promoter-CAT construct was cotransfected into C2 cells, as described previously^{17,18}, with increasing amounts of pSVE1A; total transfected DNA was kept constant (at 15 μ g), by adding appropriate amounts of carrier DNA (pUC18). CAT expression was determined using 150 μ g cell-extract-protein per assay as described^{17,18}. The MMTV promoter was induced with 10^{-6} M dexamethasone. Data is from three experiments with at least two separate plasmid preparations for each series. ▲, MMTV; ■, γ -actin; □, β -actin; ○, skeletal α -actin; ●, cardiac α -actin.

distal element either directly, or indirectly by interacting with the CARG box complexes. Constructs 6-8 do not contain the proximal CARG boxes and are slightly induced by 1 μ g pSVE1A although they become sensitive to repression at the higher pSVE1A ratio (compare 7 and 2). The basal expression of the pA10CAT plasmid was not affected by E1A coexpression, neither was the high level expression of the dexamethasone-inducible MMTV promoter.

Figure 4b shows that the pattern of pSVE1A inhibition of the cardiac α -actin promoter-constructs is similar to that of the skeletal α -actin promoter-constructs. Again there are four elements with CARG core consensus sequences; however, the two upstream boxes appear to be redundant at least for expression in the C2 cells^{18,21}. The promoter activity of all constructs that contain the proximal CARG box elements (constructs 1-4) was strongly inhibited by pSVE1A. The expression of construct 5 was unaffected by E1A cotransfection; however, the promoter activity of this fragment is very weak, directing only two to threefold induction of the basal pSV0CAT vector, and the expression is not muscle-specific¹⁸. The α -cardiac actin pA10CAT constructs (6-8) also express poorly in C2 cells¹⁸ (for example the complete fragment 6 showed ~10% of the activity of the whole promoter fragment 1). Fragment 6 functions equally well in both orientations and was sensitive to pSVE1A, although inhibition in the sense orientation was only slight with 1 μ g pSVE1A. Similarly the proximal promoter region containing all four CARG boxes (fragment 8), was only markedly sensitive to E1A at the higher pSVE1A concentration. The distal fragment 7 was not affected by pSVE1A. The SV40 promoter (in SV2CAT) was less sensitive to pSVE1A in C2 cells than its reported sensitivity in HeLa cells⁵. In summary, all constructs of the skeletal and cardiac α -actin promoters which contain the proximal CARG box elements were sensitive to a repressive effect of cotransfected E1A genes in proliferating C2 myoblasts. This repression was equivalent to the deletion or mutation of these

Fig. 4 Transient expression from α -skeletal and α -cardiac actin promoter fragments cotransfected with pSVE1A. Schematic maps of the skeletal (a) and cardiac (b) α -actin promoter regions. Positions of endogenous CARG (▲) and TATA (●) boxes are indicated. The adjoining arrows indicate promoter fragments which were fused individually to the CAT gene to create the expression plasmids as numbered. The fragments shown below the cardiac α -actin map (1-5), represent a series of 5' deletions of this promoter linked at +68 to the CAT gene in pSV0CAT. Similarly the skeletal α -actin fragment 1 is the intact promoter of this gene fused at +239 to pSV0CAT. The fragments above the two maps, α -skeletal fragments 2-8, and α -cardiac fragments 6-8, have various promoter segments cut 5' to the endogenous TATA box (except α -skeletal fragment 3 which includes the endogenous TATA box), at -87, and -47 in the α -skeletal and α -cardiac promoters respectively, and these were fused to the SV40 early promoter in pA10CAT which has a TATA box. The cloning and expression of these plasmids have been described previously^{18,19,21,22}. Fragment lengths were as follows: Skeletal (1), -2,000 to +239, (2), -2,000 to -87, (3), -2,000 to -36, (4), -800 to -90, (5), -153 to -87, (6), -2,000 to -1,300, (7), -1,300 to -710, (8), -623 to -153; Cardiac (1), -485 to +68, (2), -390 to +68, (3), -177 to -68, (4), -118 to +68, (5), -47 to +68, (6), -440 to -47, (7), -440 to -240, (8), -240 to -47. Below each schematic promoter, autoradiograms of CAT assay: TLC plates show the level of expression of the CAT gene directed by each of the promoter fragment constructs after transfection of 5 μ g of the plasmid into proliferating C2 cells (as described in Fig. 3), with 0, 1, or 5 μ g cotransfected pSVE1A. The number above each triplet assay refers to the construct number indicated in the maps. All assays have been repeated at least three times with at least two different plasmid preparations and gave essentially consistent results. There was some variability in the CAT expression between experiments which appeared to be related to the passage number of the C2 cells. In cases where constructs expressed poorly, additional cell extract was used in the CAT assay to enhance the signal, and in the case of cardiac actin construct 5, the normal assay incubation time of two hours was extended to eight hours because of the very low expression of this construct. Protein contents of the CAT assays were as follows: α -skeletal constructs 1-3, 150 μ g; 4-8, 300 μ g; α -cardiac 1-3 and 6, 150 μ g; 4, 5, 7 and 8, 600 μ g; pA10CAT, SV40 and MMTV, 150 μ g.



muscle-specific CARG elements^{18,19,21,22}. Additional upstream elements may also be sensitive to the E1A proteins.

Previous studies on myogenesis have demonstrated that biochemical differentiation may proceed when cell fusion is inhibited¹⁵⁻¹⁷. Myotube formation and irreversible commitment to terminal differentiation are not, therefore, prerequisites for muscle-specific gene expression. Biochemical differentiation, however, always accompanies myoblast fusion. This suggests that the fusion steps are controlled by multiple factors, some of which are common to the triggering of biochemical differentiation. Two classes of inhibitors may prevent myoblast fusion and biochemical differentiation: those which affect the cell cycle of the myoblast and prevent entry into G₀, and those which do not prevent exit from the cell cycle, but inhibit muscle specific gene expression by other mechanisms. Examples of the former are high mitogen or growth factor levels and some oncogenic viruses²³. Examples of the latter are transforming growth factor- β ²⁴, the activated *ras* oncogene²⁵, sodium butyrate, calcium chelators¹⁷ and metalloprotease inhibitors²⁶. The specific molecular events which cause repression of muscle-specific gene expression by either group of inhibitors is not clear. In this paper we have shown that E1A gene products cause an almost complete suppression of six endogenous muscle-specific transcripts and that, for the skeletal and cardiac α -actin genes, this inhibition is at the level of transcriptional control. Our results indicate that E1A products suppress the muscle-specific transcriptional activation mediated by the 5' CARG box elements in α -actin genes, although it is not clear whether CARG box flanking regions are important, or at what level the inhibition occurs. As the complete myogenic pathway was inhibited by E1A gene expression, it is probable that the activities of other muscle-specific promoters, possibly those involved in the initial stages of muscle cell commitment and cell fusion, are also sensitive to E1 gene products. Indeed it is possible that E1A gene expression in specialized cell-types may have a broad effect on the activities of tissue specific enhancer/promoter elements.

The molecular mechanisms of E1A-mediated transcriptional activation or repression are unknown. By *in vitro* assay we detect no differences between the binding of nuclear proteins to α actin CARG elements from wild-type myotube and from Ad DNA-transformed myoblasts (unpublished results). Similar observations have been made in the case of E1A-activated promoters, although this is currently controversial²⁷⁻²⁹. It will be interesting to determine the effects of E1A transformation upon binding of protein to CARG boxes *in vivo*.

We thank Ed Ziff for providing the plasmids pSVE1A, pN20 and pF12, Peter Evans for technical expertise and Peter Gunning for discussion. This work was supported by grants to L.K. from the NIH, Veterans Administration and the Muscular Dystrophy Association.

Received 7 October 1987; accepted 15 February 1988.

- Berk, A. J. A. *Rev. Genet.* **20**, 45-79 (1986).
- Kingston, R. E., Baldwin, A. S. & Sharp, P. A. *Cell* **41**, 3-5 (1985).
- Moran, E. & Matthews, M. B. *Cell* **48**, 177-178 (1987).
- Lillie, J. W., Green, M. & Green, M. R. *Cell* **46**, 1043-1051 (1986).
- Velcich, A. & Ziff, E. *Cell* **40**, 705-716 (1985).
- Hen, R., Borrelli, E. & Chambon, P. *Science* **230**, 1391-1394 (1985).
- Hen, R., Borrelli, E., Fromental, C., Sassone-Corsi, P. & Chambon, P. *Nature* **321**, 249-251 (1986).
- Velcich, A., Kern, F. G., Basilico, E. B. *Molec. cell Biol.* **6**, 4019-4025 (1986).
- Stein, R. W. & Ziff, E. B. *Molec. cell Biol.* **7**, 1164-1170 (1987).
- Simon, M. C. *et al. Molec. cell Biol.* **7**, 2884-2890 (1987).
- Zerler, B., Roberts, R. J., Matthews, M. B. & Moran, E. *Molec. cell Biol.* **7**, 622-639 (1987).
- Hoeffler, W. K. & Roeder, R. G. *Cell* **41**, 955-963 (1985).
- Wu, L., Rosser, D. S. E., Schmidt, M. C. & Berk, A. *Nature* **326**, 512-515 (1987).
- Smith, D. H., Kegl, D. M. & Ziff, E. D. *Molec. cell Biol.* **5**, 2684-2696 (1985).
- Wakeman, M. J. O. *Biochem. J.* **228**, 1-12 (1985).
- Entwistle, A., Curtis, D. H. & Zelin, J. J. *Cell Biol.* **103**, 857-866 (1986).
- Endo, T. & Nadal-Ginard, B. *Cell* **49**, 515-526 (1987).
- Minty, A. & Kedes, L. H. *Molec. cell Biol.* **6**, 2125-2136 (1986).
- Minty, A., Blau, H. & Kedes, L. H. *Molec. cell Biol.* **6**, 2137-2148 (1986).
- Mayer, Y., Czosnek, H., Zeelon, P. E., Yaffe, D. & Nudel, U. *Nucleic Acids Res.* **12**, 1087-1100 (1984).
- Miwa, T. & Kedes, L. *Molec. cell Biol.* **8**, 2803-2813 (1987).
- Muscato, G. & Kedes, L. *Molec. cell Biol.* **7**, 4089-4099 (1987).
- Falcone, G., Tato, F. & Alema, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 426-430 (1985).

- Massague, J., Cheifetz, S., Endo, T. & Nadal-Ginard, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8206-8210 (1986).
- Olson, E. N., Spizz, G. & Tainsky, M. A. *Molec. cell Biol.* **6**, 2104-2111 (1987).
- Baldwin, E. & Kaylar, C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8029-8033 (1986).
- Yee, A. S., Reichel, R., Kovacs, I. & Nevins, J. R. *EMBO J.* **6**, 2061-2068 (1987).
- SivaRaman, L., Subramanian, S. & Thimmappaya, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5914-5918 (1986).
- Devaux, B., Glenn, A. & Keding, C. *Molec. cell Biol.* **7**, 4560-4563 (1987).
- Babis, L. E., Young, C. S. H., Fisher, P. B. & Ginsberg, H. S. *J. Virol.* **40**, 454-465 (1983).
- Gunning, P. *et al. Molec. cell Biol.* **7**, 4100-4114 (1987).

Enhancer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription *in vitro*

Nicolas Mermod, Trevor J. Williams & Robert Tjian

Department of Biochemistry and Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA

The simian virus 40 (SV40) transcriptional enhancer is composed of multiple *cis*-acting DNA sequence motifs, each individually having a two- to fourfold effect on the efficiency of transcription¹⁻⁸. When various distinct *cis*-elements act in combination, however, a dramatic enhancement of transcription initiation often results⁹⁻¹¹. SV40-enhancer A-domain sequences were previously shown to be important for early and late transcription *in vivo*. Here we report the isolation of the enhancer binding factor AP-4, which recognizes a motif in this domain. Purified AP-4 activates SV40 late transcription *in vitro*, and this stimulation is augmented by the addition of transcription factor AP-1 which binds to adjacent sequences in the A-domain¹², suggesting coordinate action of the two factors for transcriptional enhancement. AP-1 also represses late transcription from a major *in vitro* start site which is poorly used *in vivo*, indicating that AP-1 can act as both a positive and negative regulator of SV40 late transcription. Thus by manipulating the levels of different *trans*-acting factors *in vitro*, we can recreate the pattern of SV40 late initiation observed during the viral lytic cycle *in vivo*.

The SV40 enhancer, consisting of the 72-base pair (bp) repeats and the adjacent A-domain sequences (Fig. 1), is composed of several functional *cis*-elements that are responsible for activating both early and late transcription *in vivo*¹⁻⁸. Recently, DNA binding factors that recognize motifs present in the enhancer have been detected¹³⁻¹⁸ and several of these which interact with the 72-bp repeats have been purified^{12,19-23}. One of these, AP-2, has been shown to activate early transcription *in vitro* in association with the promoter binding factor Sp1 (refs 22, 24). In contrast, the cellular factors involved in the control of SV40 late transcription are poorly characterized. Here we have focused on the factors that interact with the enhancer A-domain which is proximal to the SV40 late start-sites, and has been shown to be responsible in part for regulating SV40 late transcription *in vivo*⁸.

To analyse the factors that bind to the A-domain, we used DNA affinity chromatography to purify HeLa cell factors interacting specifically with this region of the enhancer²⁵. Figure 1 shows that a protein of relative molecular mass (*M_r*) 48,000 (48K), designated AP-4, binds selectively to the A-domain sequences of the SV40 enhancer. To ascertain whether the 48K protein was responsible and sufficient for the binding activity, this protein and a 116K polypeptide contaminant were separated by SDS PAGE, eluted and subsequently renatured to restore DNA binding activity²⁴. Binding experiments performed with both fractions showed that the 48K protein was responsible for the specific binding activity. Interestingly, in addition to interacting strongly with the enhancer, AP-4 also weakly recognizes sequences present in the SV40 21-bp repeats. Weak binding to the SV40 early promoter has also been observed for several

Fig. 1 Purification of AP-4 and localization of DNA binding sites in SV40. The preparation of HeLa cells nuclear extracts and fractionation by S-300 gel filtration chromatography were as described²⁴. The S-300 active fractions were further purified by two successive passes over sepharose resin coupled to synthetic AP-4 binding site-containing oligonucleotides²⁵. First and second passes over the affinity columns were performed using resin coupled to oligomerized (GATCAGCTGCAGCCCGC) and (GATCACAGCTGTGGAAT) sequences respectively (see also Fig. 4). This resulted in approximately 40,000-fold purification, as measured by the DNase I protection assay, and yielded 2–8 µg purified protein from 80 g cells. The 48K protein and a minor 116K contaminant were further purified by SDS-PAGE, eluted and renatured²⁴. *a*, DNase I protection analysis of AP-4 binding to SV40 intergenic control region. DNase footprinting reactions were as described¹², in the presence of ~5 ng affinity purified protein or with SDS-polyacrylamide gel-eluted and renatured 48K or 116K polypeptides. The (–) indicates control DNase I reactions performed in the absence of protein. The locations of the early initiation site and A/T rich box, 21-bp repeats (21), one copy of the 72 base pair repeat (72), enhancer A-domain, and AP-4 protected sequences are indicated. *b*, Silver-stained 8% SDS-polyacrylamide gel of proteins during AP-4 purification. 10 µl S-300 (0.2 U; 1 U corresponds to the amount of protein necessary to fully protect the SV40 A-domain sequences in a footprinting assay) first affinity column pass (2 U), second pass (10 U), or 50 µl (20 U) gel-purified and renatured 48K protein (AP-4) were loaded in parallel with molecular weight markers M; (M_r s 29, 43, 66, 97, 116 and 205K).

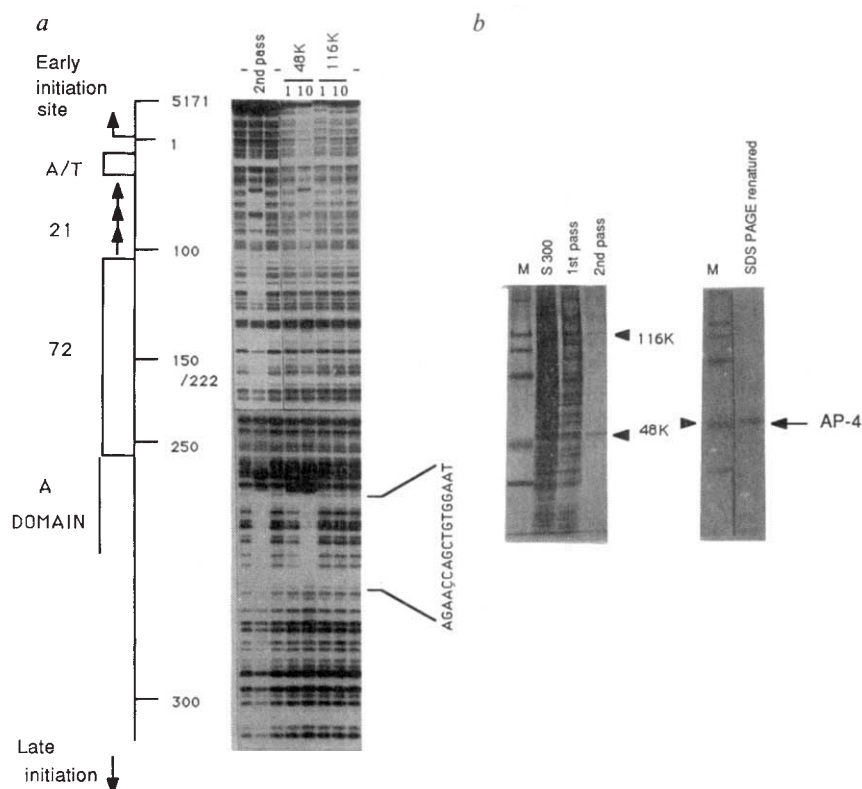
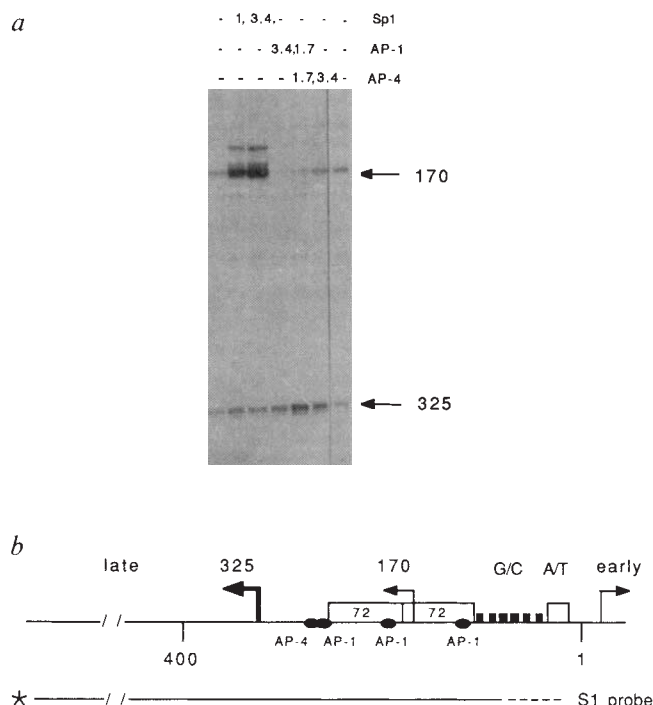


Fig. 2 Regulation of SV40 late promoter *in vitro* transcription by Sp1, AP-1 and AP-4. *a*, Transcription reactions used the Sp2 (general transcription factors) and the RNA polymerase II-containing CL225 fractions²⁸ in the presence of 100 ng pLUM DNA template, with the indicated amounts (µl) of Sp1 (1 U µl⁻¹ on SV40 G/C boxes), AP-1 and AP-4 (both 1 U µl⁻¹ as measured on SV40 enhancer A-domain). *b*, Transcripts were detected by S1 nuclease assay, using an SV40 *Ava*II-D single-stranded probe³⁰, according to ref. 39. Protected probes corresponding to the nt 170 and nt 325 initiation sites were resolved on a 4% polyacrylamide, 7M urea sequencing gel. pLUM consists of an SV40 *Eco*RI (*Eco*RII; 5092) to *Eco*RV (768) DNA fragment inserted in pxf3 (ref. 40). Filled boxes indicate the G/C boxes recognized by Sp1, present in the 21-bp repeats (see also Fig. 1 legend).



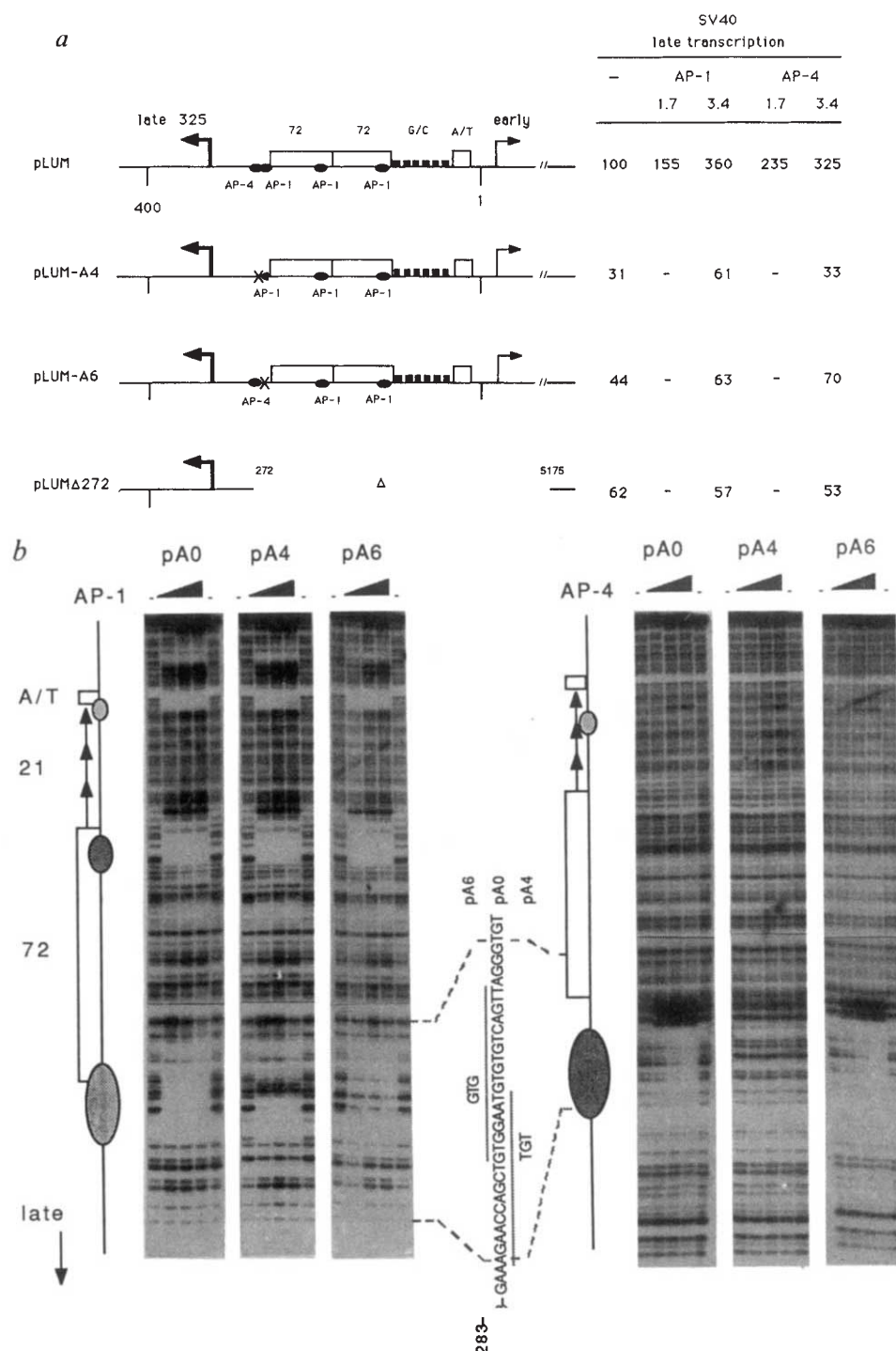
561

other factors that bind with high affinity to the 72-bp component of the enhancer (AP-1, AP-2 and AP-3, refs 12, 22). This suggests a functional mechanism by which proteins contacting the two domains could interact to stimulate transcription^{26,27}.

To investigate the role of AP-4 in SV40 late transcription, we tested the ability of this protein to activate transcription in a system reconstituted *in vitro*. Crude nuclear or whole cell transcription extracts already contain significant amounts of enhan-

cer binding factors. Therefore we used a partially purified *in vitro* transcription system consisting of fractions containing RNA polymerase II (CL225) and general transcription factors (Sp2) (ref. 28), which are mostly depleted of enhancer binding factors. SV40 late transcription normally initiates at multiple sites both *in vivo* and *in vitro*. Most transcription initiation *in vivo* occurs at nucleotide (nt) 325 (ref. 29), whereas *in vitro* an initiation site at nt 170 is used to a similar or greater extent (refs

Fig. 3 AP-1 and AP-4 regulation of *in vitro* SV40 late transcription and binding to wild-type and mutant sequences of enhancer A-domain. *a*, Transcription reactions were as in Fig. 2 using the templates and amounts (in footprinting units) of AP-1 and AP-4 shown. The extent of initiation at the main SV40 late site (nt 325) was estimated by scanning densitometry of autoradiographic exposures. Values were normalized to the basal level of transcription of the wild-type template (pLUM). (—), Not determined. The enhancer A-domain A4 and A6 mutations⁶ were introduced into the pLUM template by exchanging the *Kpn*I-*Sph*I enhancer fragment of pA4 and pA6 for the corresponding fragment of pLUM, to generate pLUM-A4 and pLUM-A6 respectively. The location of mutated sequences on the transcription templates are indicated by X. The pLUM constructs differ from the pA series by having two copies of the wild-type 72-bp repeat sequences and by lacking the introduced *Bam*HI site between the 72 and 21-bp repeats. Sequences between the *Pvu*II (nt 272) and Klenow-repaired *Hin*III (5,175) sites were deleted in pLUMΔ272. *b*, DNaseI protection experiments were performed in the absence (—) or presence of increasing amounts of AP-1 (10–80 ng) or AP-4 (1–12 ng) after two passes over the affinity resins (see also Fig. 1 legend). Footprinting probes were derived either from the pA0 wild-type or pA4 and pA6 mutant plasmids⁶. The protected sequences of the AP-1 and of AP-4 sites are respectively indicated above and below the sequence in the centre of the diagram, along with the positions of the pA4 and pA6 triple point mutations.



30, 31 and Fig. 2). The addition of purified AP-4 to our *in vitro* transcription system leads to specific stimulation of initiation from the nt 325 start site (Figs 2 and 3*a*). In contrast, transcription in the presence of Sp1, the transcription factor that binds to the 21-bp repeats, predominantly activates the cryptic nt 170 start site (refs 28, 32 and Fig. 2). These reconstitution experiments establish the function of AP-4 as a transcription factor which selectively activates initiation from the late nt 325 start site *in vitro*.

Previous data identified a weak binding site for the transcription factor AP-1 adjacent to the AP-4 site in the A-domain (see ref. 12 and Fig. 3*b*). We therefore tested the ability of AP-1 to activate late transcription *in vitro* in the presence or absence of AP-4 (Figs 2 and 3*a*). In a similar manner to AP-4, AP-1 stimulates transcription from nt 325 by ~2–3-fold *in vitro*. These

results are consistent with the two- to fourfold decrease of early or late transcription *in vivo* when A-domain recognition sequences are altered. Interestingly, the addition of half maximal amounts of both AP-1 and AP-4 together resulted in a higher level of induction than a maximal amount of either factor alone, giving a three- to fourfold induction ratio. These results suggest that a synergistic effect of both factors is operating in the cell free transcription reaction. Moreover, addition of AP-1 to the *in vitro* transcription reaction greatly reduced initiation at nt 170. Thus AP-1 has both a positive and negative effect on SV40 late transcription *in vitro*. The pattern of expression *in vitro* with high levels of AP-1 and low levels of Sp1, where the nt 325 start site predominates over the nt 170 site, is a remarkably faithful reflection of the *in vivo* transcription initiation specificity. In fact recent data suggest that an analogous situation could occur

in vivo during late viral infection, where alterations in the expression and/or activity of factors recognizing the AP-1 and Sp1 binding sites have been proposed³³⁻³⁵.

To assess whether the observed transcriptional activation caused by AP-1 and AP-4 was dependent on specific DNA sequence interactions within the A-domain, we tested the effect of various mutant templates on *in vitro* transcription. Two triple point mutations, pA4 and pA6, which decrease the A-domain enhancer activity *in vivo*⁶, were transferred to the wild-type pLUM plasmid (Fig. 3a). These two clustered point mutants, as well as a template deleted of the whole enhancer (pLUMΔ272), were then tested for their ability to activate late transcription *in vitro*. Enhancement of transcription by either AP-1 or AP-4 using the pLUMΔ272 template is essentially abolished, indicating that these factors are exerting their effect through the enhancer/promoter sequences of SV40. Both A-domain mutants pLUM-A4 and pLUM-A6 also decreased AP-1 dependent activation of transcription from the nt 325 initiation site. Similarly, pLUM-A4 transcription was not activated by AP-4, whereas some induction of transcription was observed using the pLUMA6 template (Fig. 3a). Addition of half-maximal amounts of both factors to the mutant template transcription reactions did not result in any significant activation of transcription (data not shown). These results demonstrate that the A-domain sequences are indeed important for both AP-1 and AP-4 induction of SV40 late transcription *in vitro*.

To correlate the *in vitro* transcription data with the effect of A-domain mutations upon AP-1 and AP-4 binding, the pA4 and pA6 sequences were compared to the wild-type sequence (pAO) by DNase I protection analysis. The pA6 mutation decreased AP-1 binding affinity severalfold, whereas an altered binding pattern was observed using pA4 (Fig. 3b). Thus the decreased induction of transcription obtained with AP-1 is consistent with its altered binding to the A-domain. Some residual transcriptional activation of the mutant templates is probably the result of AP-1 interactions with sequences present in the 72-bp repeats. In the case of AP-4, the pA4 mutation abolished binding of this factor to the A-domain, which strongly correlates with the loss of transcriptional induction of pLUM-A4. But the pA6 mutant showed an unaltered pattern of protection by AP-4, which is consistent with the finding that AP-4 still shows some induction of pLUM-A6 transcription. These results indicate that AP-1 and AP-4 are important in late transcription *in vitro* by binding selectively to adjacent recognition sites in the A-domain.

The basal level of transcription of the pLUM-A4 and pLUM-A6 mutants is reduced relative to the wild-type template (Fig. 3a). This is probably due to the presence of trace amounts of AP-1 and AP-4 in the Sp2 and RNA polymerase II-containing fractions respectively (data not shown). Note also that the AP-4 stimulation obtained with pLUM-A6 was much lower than using the wild-type template, even though AP-4 binding was apparently unaffected. Because the pA6 mutation disrupts AP-1 binding, a possible explanation for the lower levels of AP-4 induction is that the amounts of AP-1 contaminating the transcription extracts cannot interact with the mutated A-domain. This explanation is consistent with the proposed synergistic interaction of the two factors on the SV40 late transcription *in vitro*.

To determine whether the apparent functional association of AP-1 and AP-4 is a more general phenomenon, not just limited to the SV40 A-domain, we have tested the binding of these factors to additional enhancer/promoter sequences. AP-4 was found to bind sequences in the polyoma³⁶ virus enhancer, as well as to several cellular promoters, including the human metallothionein IIA basal level enhancer-1 (ref. 37) and the cyclic AMP-responsive element of the human proenkephalin gene³⁸ (Fig. 4). Several mutations that affect the cAMP response of the proenkephalin gene are also defective for AP-4 binding, which has suggested a role for AP-4 in the regulation of this gene, in combination with other factors (M. Comb *et al.*, manuscript in

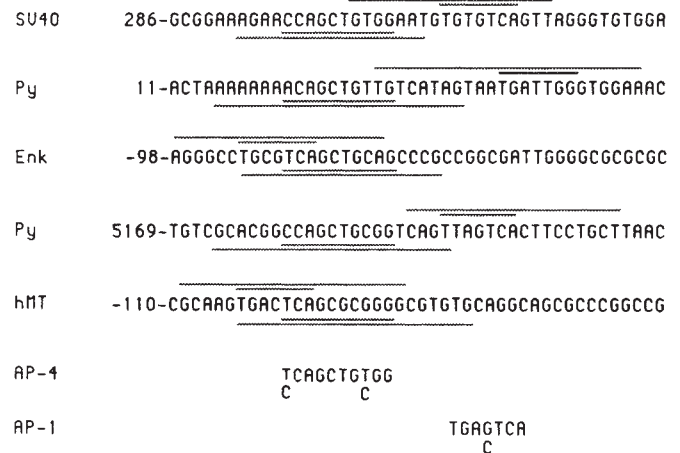


Fig. 4 Alignment of AP-4 recognition sequences on SV40 and polyoma (Py)³⁶ enhancers, human proenkephalin cAMP-responsive elements (enk)³⁸ and human metallothionein-IIA basal level enhancer element-1 (hMT)³⁷. Sequences protected by AP-4 and AP-1 are marked below and above the sequence respectively. The double lines indicate conserved nucleotides in the regions protected by AP-1 and AP-4. Binding sequences are listed in decreasing order of affinity for AP-4. Conserved nucleotides between the different binding sites are indicated (AP-4), and the AP-1 binding site consensus sequence is also shown for comparison¹². The two binding sequences shown in the polyoma enhancer are the only two AP-4 sites present either in the A3 wild-type or in the 441 mutant strains³⁶.

preparation). Strikingly, in all cases analysed so far, AP-4 recognized sequences that overlap with AP-1 binding sites (Fig. 4). This suggests that AP-1 and AP-4 may indeed function together to activate transcription from a number of different promoters. Consistent with this hypothesis, we have recently found that *in vitro* transcription of the metallothionein IIA gene in the presence of both AP-1 and AP-4 revealed a synergistic activation similar to that observed for SV40 late transcription (unpublished data).

The apparent synergism of AP-1 and AP-4 could result from the cooperative binding of the two factors to the enhancer. We do not favour this interpretation however, because preliminary binding studies have suggested that the affinity of each factor for its SV40 A-domain recognition sequence is unaltered by the binding of the other factor to the adjacent site. Alternatively, synergism of transactivation may result from the cooperative recruitment of a third as yet unidentified transcription factor by both AP-1 and AP-4. In either case, results suggest a coordinate effect of two enhancer factors upon the induction of transcription *in vitro*, which may in part explain why combined DNA enhancer motifs are more potent transcriptional elements *in vivo*.

We thank P. Chambon, M. Comb and F. Fujimura for gifts of plasmids or viruses, J. Alwine, E. Fodor and W. Herr for communicating results before publication, members of the tissue culture facility for preparation of HeLa cells and other members of the Tjian lab for helpful discussion. N.M. and T.W. were supported by a fellowship from the Swiss NSF and an ICRF travelling fellowship, respectively. This work was in part supported by a grant from the NIH to R.T.

Received 18 December 1987; accepted 19 February 1988.

1. Weiher, H., König, M. & Gruss, P. *Science* **219**, 626-631 (1983).
2. Herr, W. & Gluzman, Y. *Nature* **313**, 711-714 (1985).
3. Keller, J. M. & Alwine, J. C. *Molec. cell. Biol.* **5**, 1859-1869 (1985).
4. Firk, T. A. & Subramanian, K. N. *Molec. cell. Biol.* **6**, 3667-3676 (1986).
5. Herr, W. & Clarke, J. *Cell* **45**, 461-470 (1986).
6. Zenke, M. *et al.* *EMBO J.* **5**, 387-397 (1986).
7. Ernault-Lange, M., Omilli, F., O'Reilly, D. R. & May, E. *J. Virol.* **61**, 167-176 (1987).
8. May, E., Omilli, F., Ernault-Lange, M., Zenke, M. & Chambon, P. *Nucleic Acids Res.* **15**, 2445-2461 (1987).

9. Chiu, R., Imagawa, M., Imbra, R. J., Bockoven, J. R. & Karin, M. *Nature* **329**, 648-651 (1987).
10. Ondek, B., Shepard, A. & Herr, W. *EMBO J.*, **6**, 1017-1025 (1987).
11. Schirm, S., Jiricny, J. & Schaffner, W. *Genes Dev.*, **1**, 65-74 (1987).
12. Lee, W., Mitchell, P. & Tjian, R. *Cell* **49**, 741-752 (1987).
13. Davidson, I. et al. *Nature* **323**, 544-548 (1986).
14. Wildeman, A. G. et al. *Molec. cell. Biol.* **6**, 2098-2105 (1986).
15. Bohmann, D. et al. *Nature* **325**, 268-272 (1987).
16. Lee, W., Haslinger, A., Karin, M. & Tjian, R. *Nature* **325**, 368-372 (1987).
17. Rosales, R. et al. *EMBO J.* **6**, 3015-3025 (1987).
18. Xiao, J. H. *EMBO J.* **6**, 3005-3013 (1987).
19. Angel, P. et al. *Cell* **49**, 729-739 (1987).
20. Imagawa, M., Chiu, R. & Karin, M. *Cell* **51**, 251-260 (1987).
21. Johnson, P. F., Landschulz, W. H., Graves, B. J. & McKnight, S. L. *Genes Dev.* **1**, 133-146 (1987).
22. Mitchell, P. J., Wang, C. & Tjian, R. *Cell* **50**, 847-861 (1987).
23. Sturm, R., Baumrucker, T., Franza, B. R. & Herr, W. *Genes Devel.* (in the press).
24. Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. *Science* **234**, 47-52 (1986).
25. Kadonaga, J. T. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5889-5893 (1986).
26. Brady, J., Loeken, M. R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7299-7303 (1985).
27. Takahashi, K. et al. *Nature* **319**, 121-126 (1986).
28. Dynan, W. S. & Tjian, R. *Cell* **32**, 669-680 (1983).
29. Salzman, N. P., Venkatachala, N. & Selzer, G. B. in *The Papovaviridae*, Vol. 1, *The Polyomaviruses* (ed. Salzman N. P.) 27-98 (Plenum, New York, 1986).
30. Brady, J. N., Radonovich, M. & Salzman, N. P. *J. Virol.* **44**, 772-781 (1982).
31. Rio, D. C. & Tjian, R. *J. Molec. appl. Genet.* **2**, 423-435 (1984).
32. Dynan, W. S. & Tjian, R. *Cell* **35**, 79-87 (1983).
33. Kim, C. H., Heath, C., Bertuch, A. & Hansen, U. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6025-6029 (1987).
34. Nir, U., Fodor, E. & Rutter, W. J. *Molec. cell. Biol.* (in the press).
35. Gallo, G. J. & Alwine, J. C. *Molec. cell. Biol.* (in the press).
36. Fujimura, F. K., Deininger, P. L., Friedmann, T. & Linney, E. *Cell* **23**, 809-814 (1981).
37. Karin, M., Haslinger, A., Heguy, A., Dietlin, T. & Cooke, T. *Molec. cell. Biol.* **7**, 606-613 (1987).
38. Comb, M., Birnberg, N. C., Seasholtz, A., Herbert, E. & Goodman, H. M. *Nature* **323**, 353-356 (1986).
39. Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **7**, 1175-1193 (1979).
40. Hanahan, D. *J. molec. Biol.* **166**, 557-580 (1983).



Fig. 1 STEM images of unstained MAP 1C. A field of unfixed particles is shown, all of which show a pair of globular heads. The rest of the molecule is difficult to see at this magnification. Bar, 50 nm.

Methods. MAP 1C was prepared by ATP-extraction of taxol-stabilized calf brain microtubules and fractionated by sucrose density gradient centrifugation⁷. Peak fractions (~50 µg protein ml⁻¹) were dialysed overnight on ice against 50 mM HEPES, 50 mM PIPES, pH 7.0, containing 2.0 mM MgSO₄ and 1.0 mM EDTA. Some samples were fixed in 0.1% glutaraldehyde for 5 min at 0 °C before dialysis. The samples were adsorbed to carbon-coated electron-microscope grids and freeze-dried¹⁰. Further preparation of samples for electron microscopy and mass analysis were as described^{10,28}.

Microtubule-associated protein 1C from brain is a two-headed cytosolic dynein

Richard B. Vallee*, Joseph S. Wall†, Bryce M. Paschal* & Howard S. Shpetner*

* Cell Biology Group, Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545, USA

† Brookhaven National Laboratory, Biology Department, Upton, New York 11973, USA

Dynein, an ATPase, is the force-generating protein in cilia and flagella¹. It has long been speculated that cytoplasmic microtubules contain a related enzyme involved in cell division² or in intracellular organelle transport³. A 'cytoplasmic dynein' has been described in sea urchin eggs^{4,5}, but because the egg stockpiles precursors for both cytoplasmic and ciliary microtubules, the role of this enzyme in the cell has remained unresolved. We recently found that the microtubule-associated protein (MAP) 1C (ref. 6) from brain is a microtubule-activated ATPase⁷ that produces force in the direction corresponding to retrograde organelle transport in the cell⁸. MAP 1C has several similar properties to ciliary and flagellar dynein⁷⁻⁹. Here we show directly, using scanning transmission electron microscopy, that MAP 1C is structurally equivalent to the ciliary and flagellar enzyme and is the long-sought cytoplasmic analogue of this enzyme.

Ciliary and flagellar dyneins have been isolated as three-headed particles^{10,11} of M_r 1.95×10^6 (ref. 10) or two-headed particles^{12,13} of M_r 1.22 – 1.25×10^6 (refs 12, 14). Figures 1 and 2 show scanning transmission electron microscope (STEM) images of purified MAP 1C. Most particles, whether fixed or unfixed, clearly show two globular domains with an average diameter of 13.4 ± 1.7 nm ($n = 16$). To these are attached a pair of stalks which condense into a single thicker stalk or even a third globular region, particularly in the fixed specimens (Fig. 2). Such particles, nonetheless, are indistinguishable in mass from the more clearly two-headed particles. The most extended

Table 1 Mass of MAP 1C calculated from subunit composition

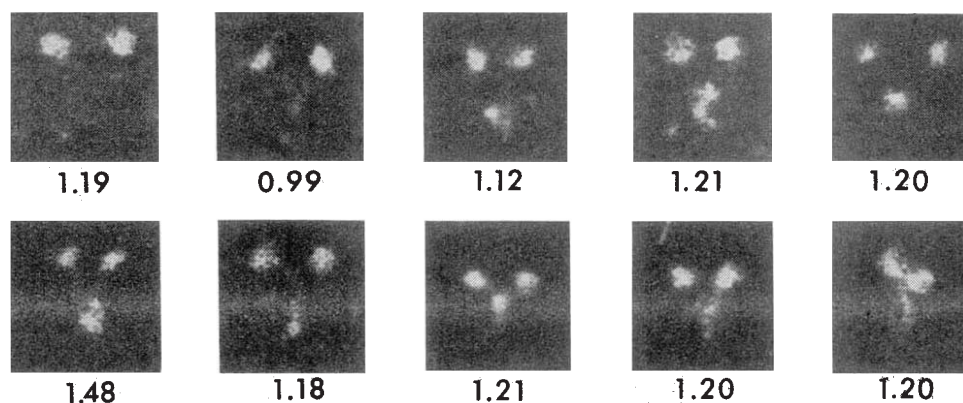
Polypeptide M_r ($\times 10^3$)	Mass ratio	Molar ratio	Proposed number of subunits	Total M_r ($\times 10^3$)
410 (MAP 1C)	1.00	2.0	2	820
74	0.28	3.1	3	222
59	0.05	0.6	1	59
57	0.05	0.7	1	57
55	0.07	1.0	1	55
53	0.08	1.2	1	53
Total				1,266

Mass ratios based on gel densitometry are from ref. 7. Molar ratios are recalculated using a M_r of 410×10^3 for the MAP 1C heavy chains, derived from the sum of the molecular mass of fragments obtained from vanadate-mediated photocleavage^{7,24,25}. In addition, we assume there are two heavy chains in MAP 1C, corresponding to the two heads observed by STEM and the two electrophoretic species seen in Fig. 4. Note that with these assumptions, the number of 74×10^3 subunits is calculated to be three, which is equal to the number of 'intermediate chains' of sea urchin flagellar and *Tetrahymena* ciliary dyneins^{16,23}. This result further emphasizes the structural similarity of MAP 1C to the axonemal dyneins, although differences in the size of the individual intermediate chains and smaller subunits indicate that MAP 1C has some distinct properties. Thus, sea urchin flagellar outer arm dynein has intermediate chains of M_r 122, 90, and 76×10^3 , and four light chains of M_r 14– 24×10^3 , in marked contrast to the MAP 1C subunits listed here.

MAP 1C particles are ~50 nm from the base to the top of the globular 'heads'. The range of images of MAP 1C, as well as the sizes of the molecules and of the individual heads, are strikingly similar to two-headed *Chlamydomonas*- and sea urchin-flagellar dyneins^{12,13}.

A compilation of mass measurements for individual MAP 1C particles is shown in Fig. 3. Peaks for both fixed and unfixed particles were centred at $M_r = 1.2 \times 10^6$. The distribution for fixed particles skewed to higher molecular mass. This probably reflects aggregation induced by the glutaraldehyde, rather than the added mass of the crosslinker itself, which is typically no

Fig. 2 STEM images of unstained MAP 1C. Methods as in Fig. 1. We chose images that depict the full range of morphology observed. The first four samples are unfixed, the rest are fixed with glutaraldehyde. Molecular masses are indicated, $\times 10^6$. Each panel is 64×64 nm.

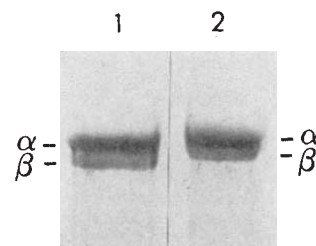


more than 5% (ref. 15). We found some evidence of dissociation or degradation in the unfixed samples, in some preparations resulting in a second, discrete peak at $\sim 1.0 \times 10^6$. Average mass values for the experiments shown in Fig. 2 are $1.14 \pm 0.16 \times 10^6$ ($n = 29$) for unfixed particles, and $1.26 \pm 0.16 \times 10^6$ ($n = 30$) for fixed particles. These values are similar to those obtained for two-headed flagellar dyneins and are lower than those for the three-headed *Tetrahymena* dynein (Fig. 3). The mass of the individual MAP 1C heads (unfixed) is $327 \pm 50 \times 10^3$ ($n = 44$), which is in the range of (but somewhat lower than) those obtained for *Chlamydomonas* flagellar and *Tetrahymena* ciliary dynein ($\sim 374 \times 10^3$ and 400×10^3 , respectively)^{10,12}.

Ciliary and flagellar dyneins contain multiple heavy-chain polypeptides corresponding to the number of heads¹⁶. We find that MAP 1C splits into two heavy-chain electrophoretic bands under conditions that have the same effect on sea urchin flagellar outer arm dynein (Fig. 4), but not under conditions that split the *Chlamydomonas* flagellar dynein heavy chain bands^{7,17}. Assuming two heavy-chain polypeptides per MAP 1C particle, the molecular mass was calculated to be 1.27×10^6 (Table 1), which fits very well with the measured mass (Fig. 3).

MAP 1C differs in mass and morphology from other high-

Fig. 4 Electrophoretic separation of two MAP 1C heavy-chain species. Electrophoresis was performed at 8 mamp cm^{-2} using a 4% polyacrylamide slab gel. The method of Laemmli²⁶ was used with the omission of SDS and the inclusion of 6 M urea in both the stacking and separating gels. The gel was stained with Coomassie brilliant blue. Only the high-molecular-mass region is shown. Lane 1, sucrose density gradient-purified MAP 1C (ref. 7); lane 2, 21S sea urchin flagellar outer-arm dynein purified by sucrose density gradient centrifugation¹³. Both mammalian and sea urchin heavy-chain bands split into two species, designated α and β following the convention used for flagellar and ciliary dyneins²⁷. The sea urchin flagellar dynein bands run slightly above those of MAP 1C.



molecular-mass microtubule-associated proteins, which are long, fibrous molecules¹⁸, and is clearly distinct from myosin¹⁹ and from kinesin^{20,21}. MAP 1C differs from the protozoan ciliary and flagellar dyneins that have been examined, which have a native three-headed structure¹⁰⁻¹², but appears to be closely comparable to two metazoan flagellar dyneins, which are two-headed^{13,14,22}. During our purification procedure, we saw no indication that MAP 1C had dissociated from a larger dynein particle⁷, as occurs with *Chlamydomonas* flagellar dynein^{11,12}, nor did we detect particles corresponding to one-headed or three-headed dyneins in the STEM analyses (Figs 1-3).

MAP 1C has force-producing and pharmacological properties consistent with an involvement in retrograde organelle transport⁸. Thus, there could be an important difference between MAP 1C and ciliary and flagellar dyneins: MAP 1C may have a membrane-binding site instead of the ATP-independent microtubule-binding site of axonemal dyneins. Our STEM analysis did not reveal any structural feature of MAP 1C that might account for this possible difference in function. But we find marked differences between MAP 1C⁷ (Table 1) and axonemal dyneins^{16,23} in the composition of lower-molecular-mass 'intermediate' and 'light chains'. We believe these are likely to specify whether the dynein molecule associates with a second microtubule or with a membranous organelle, or perhaps even with other structures, such as kinetochores.

Finally, because of our evidence that MAP 1C is structurally equivalent to flagellar and ciliary dyneins, we propose that it should now be called 'dynein'.

We thank Grace Shu, Frank Kito, Martha Simon and Paul Furcinitti for technical assistance, George Witman for extensive advice and Christine Collins and Nancy Kravit for critical reading of the manuscript.

Received 9 December 1987; accepted 23 February 1988.

1. Gibbons, I. R. *Proc. natn. Acad. Sci. U.S.A.* **50**, 1002-1010 (1963).
2. Cande, W. Z. *Nature* **295**, 700-701 (1982).
3. Beckerle, M. C. & Porter, K. R. *Nature* **295**, 701-703 (1982).
4. Weisenberg, R. C. & Taylor, E. W. *Expl Cell Res.* **53**, 372-384 (1968).

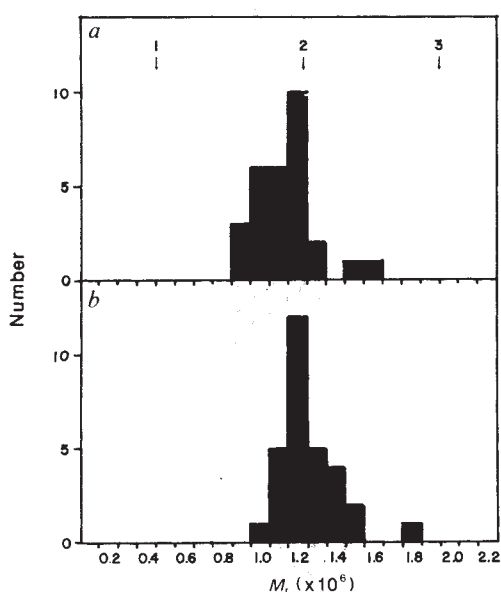


Fig. 3 Histogram of molecular mass of MAP 1C particles. Mass was determined by integration of electron scattering within a circle of 35 nm radius. Data are shown from one experiment each for fixed and unfixed material. *a*, Mass of unfixed MAP 1C, $m_{\text{avg}} = 1.14 \pm 0.16 \times 10^6$ ($n = 29$). *b*, Mass of fixed MAP 1C, $m_{\text{avg}} = 1.26 \pm 0.16 \times 10^6$ ($n = 30$). Arrows, reported molecular mass for: (1) one-headed 12S *Chlamydomonas* flagellar dynein¹²; (2) two-headed 18S *Chlamydomonas* flagellar dynein¹²; (3) three-headed *Tetrahymena* ciliary dynein¹⁰.

5. Pratt, M. M., Otter, T. & Salmon, E. D. *J. Cell Biol.* **86**, 738-745 (1980).
6. Bloom, G. S., Schoenfeld, T. A. & Vallee, R. B. *J. Cell Biol.* **98**, 320-330 (1984).
7. Paschal, B. M., Shpetner, H. S. & Vallee, R. B. *J. Cell Biol.* **105**, 1273-1282 (1987).
8. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181-183 (1987).
9. Shpetner, H. S., Paschal, B. M. & Vallee, R. B. *J. Cell Biol.* **105**, 126a (1987).
10. Johnson, K. A. & Wall, J. S. *J. Cell Biol.* **96**, 669-678 (1983).
11. Goodenough, U. & Heuser, J. *J. molec. Biol.* **180**, 1083-1118 (1984).
12. Witman, G. B., Johnson, K. A., Pfister, K. K. & Wall, J. S. *J. submicrosc. Cytol.* **15**, 193-197 (1983).
13. Sale, W. S., Goodenough, U. & Heuser, J. *J. Cell Biol.* **101**, 1400-1412 (1985).
14. Gibbons, I. R. & Fronk, E. *J. biol. Chem.* **254**, 187-196 (1979).
15. CaJacob, C. A., Frey, P. A., Hainfeld, J. F., Wall, J. S. & Yang, H. *Biochemistry* **24**, 2425-2431 (1985).
16. Johnson, K. A. *Rev. Biophys. biophys. Chem.* **14**, 161-188 (1985).
17. Pfister, K. K., Fay, R. B. & Witman, G. B. *Cell Motil.* **2**, 525-547 (1982).
18. Voter, W. A. & Erickson, H. P. *J. ultrastruct. Res.* **80**, 374-382 (1982).
19. Harrington, W. F. & Rodgers, M. E. *A. Rev. Biochem.* **53**, 35-73 (1984).
20. Bloom, G. S., Wagner, M. C., Pfister, K. K. & Brady, S. T. *Biochemistry* (in press).
21. Amos, L. A. *J. Cell Sci.* **87**, 105-111 (1987).
22. Marchese-Ragona, S. P., Gagnon, C., White, D., Belles Isles, M. & Johnson, K. A. *Cell Motil.* **8**, 368-374 (1988).
23. Bell, C. W., Fronk, E. & Gibbons, I. R. *J. supramolec. Struct.* **11**, 311-317 (1979).
24. Lee-Eiford, A., Ow, R. A. & Gibbons, I. R. *J. biol. Chem.* **261**, 2337-2342 (1986).
25. King, S. M. & Witman, G. B. *J. biol. Chem.* **262**, 17596-17604 (1987).
26. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
27. Witman, G. B. & Pfister, K. *J. biol. Chem.* **259**, 12072-12080 (1984).
28. Vallee, R. B., Paschal, B. M., Shpetner, H. S. & Wall, J. S. *J. Cell Biol.* **105**, 125a (1987).

Localization of the binding site for the human high-affinity Fc receptor on IgG

Alexander R. Duncan, Jennifer M. Woof*,
Lynda J. Partridge*, Dennis R. Burton*
& Greg Winter†

Medical Research Council Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK
and *Department of Biochemistry, University of Sheffield,
Sheffield S10 2TN, UK

A major pathway in the clearance of pathogens involves the coating of the pathogen with specific antibodies, and the binding of the antibody Fc region to cell receptors. This can trigger engulfment of the pathogen by phagocytes or lysis by killer cells. By oligonucleotide site-directed mutagenesis we have engineered a single amino acid change in a mouse IgG2b antibody (Glu 235→Leu) which now enables the antibody to bind to the FcRI (high affinity) receptor on human monocytes with a 100-fold improvement in affinity. This indicates that Leu 235 is a major determinant in the binding of antibody to FcRI and that the receptor may interact directly with the region linking the CH2 domain to the hinge. Tailoring the affinity of antibodies for cell receptors could help dissect their role in clearing pathogen.

Antibodies fulfil their protective role by binding to antigen through the variable domains. The antigen-antibody complex is then recognized through sites in the Fc domain by effector molecules including complement and cell receptors¹. Binding to cells can trigger phagocytosis² and antibody-dependent cell-mediated cytotoxicity (ADCC)³⁻⁵. Three receptors, FcRI, FcRII and FcRlo, which are mainly specific for human IgG1 and IgG3 isotypes, are expressed on a variety of human leukocytes. FcRI- and FcRII-like receptors are present in the mouse, and human FcRI binds to the mouse IgG2a isotype with high affinity. Human FcRI is always present in combination with either FcRII or FcRlo, and normally only on mononuclear phagocytes. It binds to monomeric human IgG1 and IgG3 isotypes with a high affinity, in contrast to FcRII and FcRlo, which bind to aggregated antibody (see ref. 6 for review). Each of the receptors appears to be involved in ADCC³⁻⁵.

The binding site for human FcRI lies on the Fc fragment, and is thus to the C-terminal side of residue 224 in human IgG1 (ref. 7). It appears to be located in the CH2 domain of the antibody (refs 8-10 and refs cited therein). Studies with human

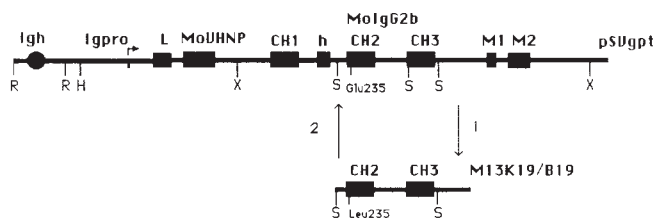


Fig. 1 Construction of mutant Glu 235→Leu mouse IgG2b antibody. Igh, IgG heavy chain enhancer; Igpro, IgG heavy chain promoter; L, exon for heavy chain signal peptide; MoUHN, exon for heavy chain variable domain; CH1-3, exons for heavy chain constant domains; h, hinge; M1 and M2, membrane domain. Restriction sites: R, *EcoRI*; H, *HindIII*; S, *SacI*; X, *XbaI*.

Methods. To make the Glu 235→Leu mutant, the vector for expression of the heavy chain¹³ was partially digested with *SacI* to release a fragment containing the CH2 and CH3 exons of the mouse $\gamma 2b$ gene. This was subcloned in the vector M13K19 (step 1), the internal *SacI* site was removed and the Glu 235→Leu mutant constructed using the synthetic mismatch primers 5'-CTTGTGGGGCTCTGA-3' and 5'-GGTCCACCCAGGAG-GTTAGG-3', respectively, in two cycles of mutagenesis²⁵. The *SacI* fragment was excised and recloned in the expression vector (step 2). After transfection into the myeloma J558L, secreted antibody was purified on a 1 ml column of NIP-Sepharose¹³.

myeloma proteins in which the hinge was deleted suggested that the top of the CH2 domain must be accessible for binding to FcRI¹¹, and sequence comparisons of several antibodies and their properties, led to a proposal that the monocyte-binding site was located mainly in the hinge link, involving residues Leu 234-Pro 238¹². In particular, it was noted that in human IgG1 or mouse IgG2a, which have a high affinity for human FcRI, the sequence in this region is Leu 234-Leu-Gly-Gly-Pro 238. However in mouse IgG2b, which does not bind human FcRI, the sequence is Leu 234-Glu-Gly-Gly-Pro 238. We have therefore altered Glu 235 to leucine in the mouse IgG2b heavy chain, and tested the mutant antibody for binding to human FcRI.

A chimaeric antibody has been described with a variable domain which binds to the hapten NIP (3-iodo-4-hydroxy-5-nitrophenylacetate) and which has constant domains of the mouse IgG2b isotype. The heavy chain is assembled in a pSVgpt vector with Ig heavy chain enhancer and promoter, and on transfection into J558L myeloma cells which carry the mouse λ light chain, antibody is secreted which can be purified on NP affinity columns¹³. The Glu 235→Leu mutant was constructed as described in Fig. 1, and ~5 mg of mutant antibody was prepared from 11 roller bottles of growing cells.

The wild-type chimaeric mouse IgG2b antibody does not inhibit the binding of pooled human IgG (mainly IgG1) to FcRI on human U 937 cells, but the Glu-235→Leu mutant inhibits as effectively as pooled human IgG (Fig. 2). Previous results in which weak binding had been detected for the IgG2b isotype¹² may have been due to trace contamination of mouse IgG2a from the ascites fluid. The direct binding of ¹²⁵I-labelled mutant to U 937 cells was quantified using Scatchard analysis, a typical plot being shown in Fig. 3. Four such experiments yielded a mean K_a value of $3.13 \pm 0.58 \times 10^8 \text{ M}^{-1}$ ($\pm$ s.e.m.) with $3.6 \pm 0.35 \times 10^3$ ($\pm$ s.e.m.) receptors per cell. The binding constant agrees closely with values for human IgG1 binding to FcRI on monocytes and U 937 cells^{12,14-16}. Thus the mutation Glu 235→Leu in mouse IgG2b antibody is sufficient to improve its affinity for FcRI by a factor of more than 100-fold, and in both the direct binding and competition experiments, it behaves as human IgG1.

Since leucine is conserved at residue 235 in antibodies which bind to FcRI¹², it is likely to be a major determinant for binding to FcRI. Other residues in the hinge link might also be determinants. Thus excision of the hinge link in the Fc fragment of human IgG1 (cleaved between 234 and 235)¹⁷ and mouse IgG2a

† To whom correspondence should be addressed.

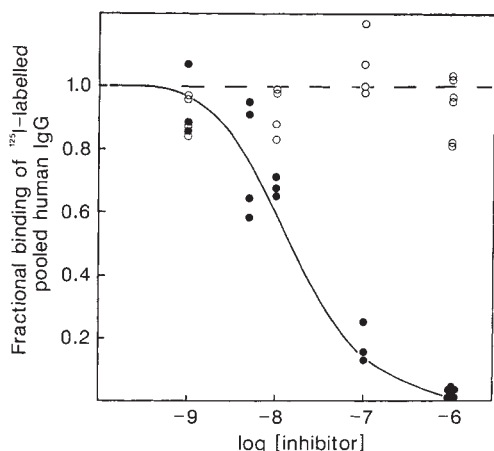


Fig. 2 Inhibition of ^{125}I -labelled pooled human IgG binding to high affinity Fc receptors (FcRI) on U937 cells by monomeric mouse IgG2b immunoglobulins. (○), Wild type IgG2b; (●), Glu 235 → Leu mutant IgG2b. For methods see Fig. 3 legend.

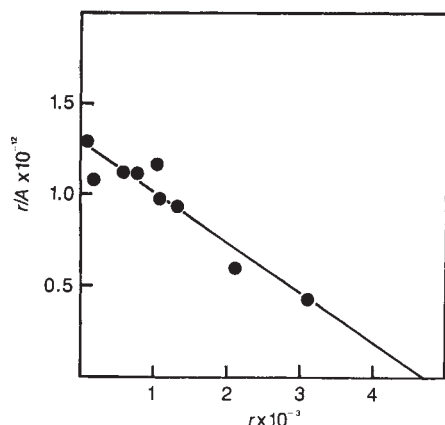


Fig. 3 Scatchard plot of ^{125}I -labelled mutant Glu235 → Leu mouse IgG2b binding to high affinity receptors (FcRI) on U937 cells. r , Number of moles of ^{125}I -(Glu 235 → Leu) mouse IgG2b antibody bound per mole of cells. A , Concentration of free ^{125}I -mutant IgG2b⁸. The number of receptors per cell is lower than those previously reported^{14,16}, but a Scatchard analysis of ^{125}I -labelled pooled human IgG binding to the U937 cells was similar (not shown). The diminished values for receptor number may be caused by growing U937 to high cell concentrations (0.9×10^6 per ml). **Methods.** The IgG-FcRI binding assay was essentially as previously described⁸, except that after introduction of water-immiscible oil to the equilibrium mixture followed by rapid centrifugation, the pelleted cells (bound ^{125}I -IgG) and medium (free ^{125}I -IgG) were separated by slicing through the tube within the oil layer.

(cleaved between 233 and 234)¹⁸ resulted in a loss of binding to human FcRI^{19,20}, although in these two cases the two CH2 domains of the antibody are no longer tethered together by the hinge disulphides. In the alignment of ref. 12, antibodies with substitutions at residues 231 and 233 still bind tightly to FcRI, but those with changes at residue 234 have a reduced affinity. Furthermore residues 236–238 are completely conserved, except in mouse IgG1 and human IgG2, which do not bind to human FcRI. Much of the link, in particular residues 234–238, may therefore be required for binding to human FcRI.

The hinge link is mobile in the crystallographic structure of human Fc²¹ and is accessible to proteolytic attack. Thus papain cleaves between residues 233 and 234 in mouse IgG2a and IgG2b¹⁸; pepsin between residues 234 and 235 in human IgG1²² and residues 238 and 239 in mouse IgG1²³; thermolysin between residues 234 and 235 in human IgG1¹⁷. The facile proteolysis

of several IgG isotypes in this region may simply reflect the underlying design of the FcRI binding site. The site appears to be accessible and flexible and would permit, for example, a hinge dislocation on binding to FcRI²⁴.

In conclusion, our results suggest that the hinge link, either as a single flexible strand or paired with the strand from the other heavy chain, is a major determinant in binding of antibody to FcRI, and we would predict that changing Leu 235 for glutamic acid (and perhaps other side chains) would destroy the interaction of human IgG1 or IgG3 with FcRI. The possibility of turning on and off the interaction of antibody with human FcRI, could help dissect the role of this receptor in phagocytosis and cell mediated lysis and in antibody therapy. Furthermore in imaging of solid tumours, eliminating interactions with FcRI could help reduce background due to antibody binding to cells with high affinity receptors in the lymphatics, liver and spleen.

We thank M. S. Neuberger for the mouse IgG2b expression vector, M. S. Neuberger and C. Milstein for advice, and M. Clark for comments on the draft. This work was supported by the Medical Research Council and the Wellcome Trust. D.R.B. is a Jenner Fellow of the Lister Institute of Preventive Medicine.

Received 13 January; accepted 3 March 1988.

- Burton, D. R. *Molec. Immun.* **22**, 161–206 (1985).
- Silverstein, S. C., Steinman, R. M. & Cohn, Z. A. *Rev. Biochem.* **46**, 669–722 (1977).
- Shen, L., Guyre, P. M. & Fanger, M. W. *J. Immun.* **139**, 534–538 (1987).
- Graziano, R. F. & Fanger, M. W. *J. Immun.* **139**, 3536–3541 (1987).
- Karpovsky, B., Titus, J. A., Stephany, D. A. & Segal, D. M. *J. exp. Med.* **160**, 1686–1701 (1984).
- Anderson, C. L. & Looney, R. J. *Immun. Today* **7**, 264–266 (1987).
- Frangione, B. & Milstein, C. *Nature* **216**, 939–941 (1967).
- Woof, J. M., Nik Jaafar, M., Jefferis, R. & Burton, D. R. *Molec. Immun.* **21**, 523–527 (1984).
- Leatherbarrow, R. J. *et al. Molec. Immun.* **22**, 407–415 (1985).
- Partridge, L. J., Woof, J. M., Jefferis, R. & Burton, D. R. *Molec. Immun.* **23**, 1365–1372 (1986).
- Klein, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 524–528 (1981).
- Woof, J. M., Partridge, L. J., Jefferis, R. & Burton, D. R. *Molec. Immun.* **23**, 319–330 (1986).
- Neuberger, M. S. & Williams, G. T. *Phil. Trans. R. Soc. A317*, 425–432 (1986).
- Anderson, C. L. & Abraham, G. N. *J. Immun.* **125**, 2735–2741 (1980).
- Kurlander, R. J. & Batker, J. *J. clin. Invest.* **69**, 1–8 (1982).
- Fries, L. F., Hall, R. P., Lawley, T. J., Crabtree, G. R. & Frank, M. M. *J. Immun.* **129**, 1041–1049 (1982).
- Hunneyball, I. M. & Stanworth, D. R. *Immunology* **30**, 579–586 (1976).
- Frankus, T. & Birshtein, B. K. *Biochemistry* **17**, 4324–4331 (1978).
- Ratcliffe, A. & Stanworth, D. R. *Immunology* **50**, 93–100 (1983).
- McCool, D., Birshtein, B. K. & Painter, R. H. *J. Immun.* **135**, 1975–1980 (1985).
- Deisenhofer, J. *Biochemistry* **20**, 2361–2370 (1981).
- Frangione, B. & Milstein, C. *J. molec. Biol.* **33**, 893–906 (1968).
- Svasti, J. & Milstein, C. *Nature* **228**, 930–935 (1970).
- Burton, D. R. *Immun. Today* **7**, 165–167 (1986).
- Carter, P., Bedouelle, H. & Winter, G. *Nucleic Acids Res.* **13**, 4431–4443 (1985).
- Raychaudhuri, G., McCool, D. & Panter, R. H. *Molec. Immun.* **22**, 1009–1019 (1985).

Dissecting the catalytic triad of a serine protease

Paul Carter & James A. Wells

Department of Biomolecular Chemistry, Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

Serine proteases are present in virtually all organisms and function both inside and outside the cell¹; they exist as two families, the 'trypsin-like' and the 'subtilisin-like', that have independently evolved a similar catalytic device² characterized by the Ser, His, Asp triad, an oxyanion binding site, and possibly other determinants that stabilize the transition state (Fig. 1)^{2–4}. For *Bacillus amyloliquefaciens* subtilisin, these functional elements impart a total rate enhancement of at least 10^9 to 10^{10} times the non-enzymatic hydrolysis of amide bonds. We have examined the catalytic importance and interplay between residues within the catalytic triad by individual or multiple replacement with alanine(s), using site-directed mutagenesis^{5,6} of the cloned *B. amyloliquefaciens* subtilisin gene⁷. Alanine substitutions were chosen to minimize unfavourable steric contacts and to avoid imposing new charge interactions or hydrogen bonds from the

Table 1 Kinetic parameters of mutant subtilisins with the substrate *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide at pH 8.60

Enzyme	Active site configuration			k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\text{s}^{-1}\text{M}^{-1}$)	$k_{\text{cat}}(\text{mutant})$
	Ser221	His64	Asp32				$k_{\text{cat}}(\text{S24C})$
Wild type	+	+	+	$(4.4 \pm 0.1) \times 10^1$	180 ± 10	$(2.5 \pm 0.1) \times 10^5$	0.74 ± 0.01
S24C	+	+	+	$(5.9 \pm 0.2) \times 10^1$	220 ± 20	$(2.7 \pm 0.2) \times 10^5$	1
S24C:S221A	—	+	+	$(3.4 \pm 0.1) \times 10^{-5}$	420 ± 40	$(8.2 \pm 0.6) \times 10^{-2}$	$(5.8 \pm 0.1) \times 10^{-7}$
S24C:H64A	+	—	+	$(3.8 \pm 0.2) \times 10^{-5}$	390 ± 50	$(9.6 \pm 1.0) \times 10^{-2}$	$(6.4 \pm 0.2) \times 10^{-7}$
S24C:D32A	+	+	—	$(2.3 \pm 0.2) \times 10^{-3}$	480 ± 80	4.7 ± 0.7	$(3.8 \pm 0.2) \times 10^{-5}$
S24C:D32A:H64A	+	—	—	$(2.6 \pm 0.1) \times 10^{-4}$	270 ± 50	$(9.4 \pm 1.6) \times 10^{-1}$	$(4.3 \pm 0.1) \times 10^{-6}$
S24C:H64A:S221A	—	—	+	$(2.8 \pm 0.2) \times 10^{-5}$	290 ± 40	$(9.6 \pm 1.3) \times 10^{-2}$	$(4.8 \pm 0.2) \times 10^{-7}$
S24C:D32A:S221A	—	+	—	$(2.8 \pm 0.1) \times 10^{-5}$	310 ± 40	$(9.2 \pm 0.9) \times 10^{-2}$	$(4.8 \pm 0.1) \times 10^{-7}$
S24C:D32A:H64A:S221A	—	—	—	$(3.0 \pm 0.1) \times 10^{-5}$	230 ± 20	$(1.3 \pm 0.1) \times 10^{-1}$	$(5.1 \pm 0.1) \times 10^{-7}$
				k_{buffer} (s^{-1})			
							$k_{\text{cat}}(\text{S24C})$
No enzyme	none			$(1.1 \pm 0.1) \times 10^{-8}$	—	—	$(1.9 \pm 0.1) \times 10^{-10}$

Mutants are abbreviated by the single-letter code for the wild-type amino acid followed by its codon position and the amino acid replacement; multiple mutants are designated by listing single mutant components separated by colons (for example, double mutant Ser221 to Cys, Ser221 to Ala is designated S24C:S221A). Construction of the mutants S24C and H64A and the double mutant S24C:H64A was as described^{12,13}. The mutations D32A and S24C were constructed simultaneously using a 48-mer oligonucleotide^{5,6} and the S221A mutant was constructed by cassette mutagenesis²⁵. The remaining multiple mutants were constructed by 3-way ligations using a 6 kb *EcoRI*/*Bam*HI fragment from the vector pSS5 (B. Cunningham, D. Powers, and J. W., unpublished) and two subtilisin fragments from appropriate mutants. Mutant constructions were verified by dideoxy sequencing²⁶. Mutant plasmids were expressed in a protease deficient strain of *B. subtilis*, BG2036²⁷. Rescue of active site mutants by co-culturing with the mutant A48E and purification was as described¹². Mutant subtilisins were assayed with the substrate, *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (Sigma). Six hydrolysis assays were performed simultaneously against substrate blanks in 1 ml 100 mM Tris-HCl (pH 8.60) 4% (v/v) dimethylsulphoxide at $(25 \pm 0.2)^\circ\text{C}$ using a Kontron Uvikon 860 spectrophotometer. Initial reaction rates were determined from the increase in absorbance at 410 nm on release of *p*-nitroaniline ($\epsilon_{410} = 8,480 \text{ M}^{-1} \text{ cm}^{-1}$)²⁸. The total substrate concentration in each assay was determined from the A_{410} after complete hydrolysis. The initial rate data were fitted to the Michaelis-Menten relationship using least squares analysis to determine K_m and V_{max} . Turnover number (k_{cat}) was calculated from the spectrophotometrically determined enzyme concentration ($\epsilon_{280}^{0.1\%} = 1.17$)²⁹. Enzyme concentrations in the assays were 30–110 $\mu\text{g ml}^{-1}$ for the active site mutants and 1 $\mu\text{g ml}^{-1}$ for the wild type and S24C enzymes. Catalytic triad residues are represented by (+) and Ala replacements by (—). Data are presented $\pm$ standard errors and the spontaneous hydrolysis rate of substrate under these conditions is shown as k_{buffer} .

substituted side chains. In contrast to the effect of mutations in residues involved in substrate binding^{8–10}, the mutations in the catalytic triad greatly reduce the turnover number and cause only minor effects on the Michaelis constant. Kinetic analyses of the multiple mutants demonstrate that the residues within the triad interact synergistically to accelerate amide bond hydrolysis by a factor of $\sim 2 \times 10^6$.

Subtilisin is synthesized as a membrane-associated precursor (preprosubtilisin)⁷. When expressed in a protease-deficient strain of *B. subtilis*, mature *B. amyloliquefaciens* subtilisin is efficiently released into the medium after autoproteolytic cleavage¹¹. Mutagenesis of the catalytic residues in subtilisin (which essentially inactivates the protease) disrupts this processing, but processing can be restored by co-culturing the mutants with a small amount of a *B. subtilis* strain (called a 'helper') harbouring an active subtilisin gene¹². We have constructed a series of active site mutants in which the catalytic triad residues are replaced by alanine in every possible combination (ref. 12, Table 1). Each mutant also contains a surface-accessible Ser24 to Cys mutation¹³ designated S24C (mutant enzymes are named using the single letter code for amino acids to indicate the substitutions made, see Table 1). The S24C substitution permits reversible attachment to an activated thiol sepharose column thereby eliminating traces of contaminating helper subtilisin which is cysteine-free¹².

The hydrolysis of the substrate (*N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide) by most of the active site mutants produced only small absorbance changes (ΔA_{410} of 0.01 to 0.10) over long periods (up to 12 h), yet the data exhibit typical Michaelis-Menten saturation behaviour (Fig. 2) with standard errors almost as small as those for wild-type subtilisin (Table 1). No detectable loss of catalytic activity occurred even during the longest kinetic runs. In addition, the background (non-enzymatic) hydrolysis of substrate was $\leq 25\%$ of the catalysed rate for even the least active enzymes (Fig. 2). The non-enzymatic

hydrolysis was subtracted directly from the enzyme assays using blank substrate solutions in a double beam spectrophotometer.

Kinetic analysis of the active site single mutants (Table 1) shows that replacement of the catalytic serine, histidine or aspartate causes a drop in turnover number (k_{cat}) by factors of 2×10^6 , 2×10^6 and 3×10^4 , respectively. The 100-fold lower values of k_{cat} which result from substitution of Ser221 and His64, compared with Asp32, are consistent with their more central role in catalysis (Fig. 1). Each mutation causes a small increase in the Michaelis constant (K_m) (~ 2 -fold) which may result from slightly altered substrate binding contacts. (Wild-type subtilisin has a two-step enzyme mechanism where deacylation is > 33 times faster than acylation¹⁴, so that K_m is a good approximation of the enzyme-substrate dissociation constant (K_s)¹⁵. As the enzyme mechanism must be changed for at least some of the mutants (see below), K_m may be less than K_s .)

Additional mutagenesis of the S24C:S221A enzyme to replace either Asp32, His64 or both, causes essentially no further change in k_{cat} or K_m (Table 1). By comparison, further mutagenesis of the S24C:D32A parent enzyme to substitute His64 or both His64 and Ser221, further reduces k_{cat} by 9 and 76-fold, respectively, with essentially no change in K_m . These data suggest that His64 provides a catalytic advantage of ~ 10 -fold to the S24C:D32A enzyme, and that Ser221 provides ~ 10 -fold advantage to the S24C:D32A:H64A enzyme. As with the S24C:S221A family of mutants, additional mutations in the S24C:H64A enzyme to replace Ser221 or both Ser221 and Asp32 do not affect k_{cat} . But replacement of Asp32 alone in the S24C:H64A mutant to give S24C:D32A:H64A, actually increases k_{cat} 7-fold. Thus, Asp32 is a liability to the S24C:H64A enzyme, possibly because of an unfavourable electrostatic effect upon catalysis (see below).

The single and multiple mutant analyses show that the catalytic effects are non-additive in two ways. First, there is a gross discrepancy between the relative drop in k_{cat} resulting from the triple alanine mutant (2×10^6 , Table 1) compared with

Table 2 Kinetic parameters of mutant subtilisins with the substrate *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide at pH 9.70

Enzyme	Active site configuration			k_{cat} (s ⁻¹)	K_{m} (μM)	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ M ⁻¹)	k_{cat} (pH 9.7)
	Ser221	His64	Asp32				k_{cat} (pH 8.6)
Wild type	+	+	+	$(6.3 \pm 0.1) \times 10^4$	440 ± 30	$(1.4 \pm 0.1) \times 10^5$	1.4 ± 0.1
S24C	+	+	+	$(8.1 \pm 0.2) \times 10^4$	560 ± 30	$(1.5 \pm 0.1) \times 10^5$	1.4 ± 0.1
S24C:S221A	—	+	+	$(5.4 \pm 0.3) \times 10^{-5}$	650 ± 90	$(8.4 \pm 1.0) \times 10^{-2}$	1.6 ± 0.1
S24C:H64A	+	—	+	$(1.9 \pm 0.1) \times 10^{-4}$	1300 ± 150	$(1.5 \pm 0.2) \times 10^{-1}$	5.1 ± 0.2
S24C:D32A	+	+	—	$(1.8 \pm 0.1) \times 10^{-2}$	1400 ± 120	$(1.3 \pm 0.1) \times 10^1$	7.8 ± 0.4
S24C:D32A:H64A	+	—	—	$(1.8 \pm 0.1) \times 10^{-3}$	460 ± 40	3.8 ± 0.3	6.9 ± 0.3
S24C:H64A:S221A	—	—	+	$(5.2 \pm 0.2) \times 10^{-5}$	480 ± 60	$(1.1 \pm 0.1) \times 10^{-1}$	1.9 ± 0.1
S24C:D32A:S221A	—	+	—	$(5.9 \pm 0.3) \times 10^{-5}$	460 ± 80	$(1.3 \pm 0.2) \times 10^{-1}$	2.1 ± 0.1
S24C:D32A:H64A:S221A	—	—	—	$(7.8 \pm 0.3) \times 10^{-5}$	730 ± 70	$(1.1 \pm 0.1) \times 10^{-1}$	2.6 ± 0.1
				k_{buffer} (s ⁻¹)	k_{buffer} (pH 9.7)		
					k_{buffer} (pH 8.6)		
No enzyme	none			$(2.8 \pm 0.1) \times 10^{-8}$	—	—	2.5 ± 0.1

Kinetic data were determined as for Table 1 except that 100 mM 3-[cyclohexylamino]-2-hydroxyl-1-propane buffer (pH 9.70) was used. Ionic strength was normalized with NaCl.

the product of the relative effects from the three single alanine mutants ($\sim 10^{17}$). Second, the double alanine mutants that retain singly the catalytic Ser, His or Asp are only a factor of 8, 0.9 or 0.9 larger in k_{cat} , respectively, than the triple alanine mutant. The product of these values (~ 6) is much below the relative k_{cat} value of 2×10^6 for wild type (S24C) compared with the triple alanine mutant. Thus, non-additive effects are shown either by subtraction of catalytic residues relative to wild-type enzyme or by addition of single catalytic residues relative to the triple alanine mutant.

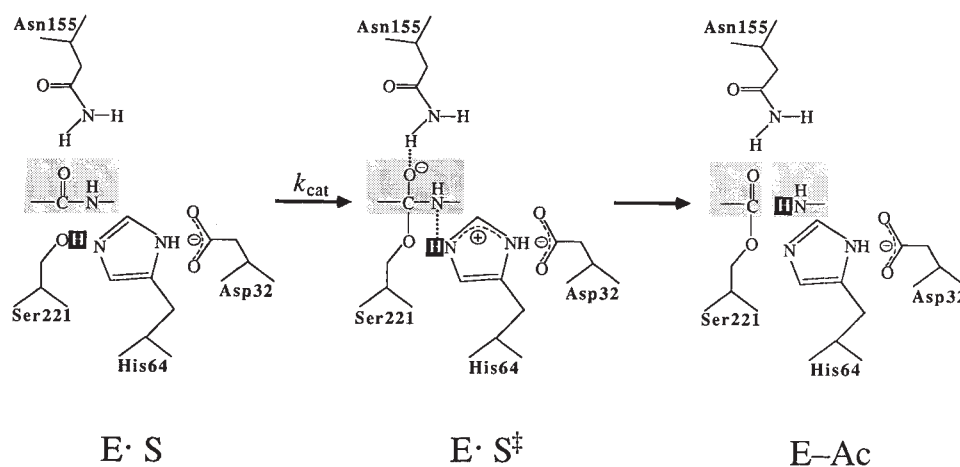
Replacement of residues in the catalytic triad with alanines necessarily perturbs the enzyme mechanism. In particular, it has been observed that in the absence of the catalytic His64 in subtilisin¹² or the catalytic Asp102 in trypsin^{16,17}, there is a marked increase in the hydroxide dependence of catalysis between pH 8 and 10 compared to the wild-type enzymes. Comparisons of the kinetic parameters for all of the catalytic triad mutants at pH 9.70 and pH 8.60 (Table 2) show that those retaining Ser221 have a substantially stronger pH dependence of k_{cat} (increased 5- to 8-fold) than enzymes containing an intact catalytic triad (increased 1.4-fold), or enzymes lacking Ser221 (increased 1.6- to 2.6-fold), or when compared with the non-enzymatic rate (increased 2.5-fold). For all enzymes the K_m values at pH 9.70 are increased between 1.5 and 3.3-fold. Preliminary evidence suggests that this effect upon K_m may result (at least partially) from ionization of Tyr104, resulting in electrostatic repulsion of the P5 succinyl group (see Fig. 3, and D.

Estell, T. Graycar, D. Powers and J. A. Wells, unpublished results).

For mutants that retain Ser221, the simplest interpretation of the data is that they continue to use Ser221 as the catalytic nucleophile. The presence of Ser221 provides a catalytic advantage of ~ 10 -fold to the S24C:D32A:H64A enzyme and ~ 100 -fold to the S24C:D32A enzyme. Furthermore, replacing His64 in the S24C:D32A enzyme causes k_{cat} to drop ~ 10 -fold, suggesting that His64 functions here to some extent (presumably as a proton acceptor for the nucleophilic Ser221). In addition, if deprotonation of the Ser221 hydroxyl is a prerequisite for nucleophilic attack in these mutants, then it is reasonable for k_{cat} to depend on hydroxide ion concentration, as observed (Table 2). Finally, in the absence of His64, the catalytic aspartate should inhibit deprotonation of Ser221 and have a deleterious electrostatic effect upon k_{cat} , as indeed was found (k_{cat} for S24C:H64A is 10-fold lower than the k_{cat} for S24C:D32A:H64A in Table 1). Like wild-type subtilisin, we anticipate the S221A family of enzymes should have a two-step enzyme mechanism. For these mutants, if deacylation is rate-determining, it is possible that the K_m values are substantially less than the K_s values¹⁵.

For the S24C:S221A family of enzymes, the reaction cannot proceed by the usual serine acyl-enzyme intermediate. Instead, direct attack of water on the scissile peptide bond may occur to produce a single tetrahedral intermediate that collapses to give the hydrolysed products. Nucleophilic attack by water is

Fig. 1 Schematic diagram showing the rate limiting acylation step in the hydrolysis of peptide bonds by subtilisin. In going from the Michaelis enzyme-substrate complex ($E \cdot S$) to the transition state complex ($E \cdot S^\ddagger$), the proton on Ser221 (darkly shaded) is transferred to His 64, thus permitting nucleophilic attack on the scissile peptide bond²⁻⁴. The proton is then transferred to the amine leaving group to generate the acyl-enzyme intermediate ($E\text{-Ac}$) (as for Asp102 in trypsin^{2-4,16,17}) is believed to position the correct tautomer of His64 for catalysis in the $E \cdot S$ complex and stabilize the protonated form of His64 in the $E \cdot S^\ddagger$ complex. Some of the hydrogen bonds that form in the $E \cdot S^\ddagger$ complex are shown by dotted lines. In deacylation these steps are reversed and water (as the nucleophile) replaces the amine leaving group.



consistent with the weak hydroxide dependence of k_{cat} for the S24C:S221A-containing mutants. The lack of a deleterious electrostatic effect from Asp32 is also consistent with a neutral attacking nucleophile (compare S24C:H64A:S221A with S24C:D32A:H64A:S221A in Table 1). It is unlikely that the S221A group of enzymes use the other members of the catalytic triad because there is no additional kinetic advantage for including the His64 or Asp32. (Strictly, we cannot be sure that the residual members of the triad are catalytically inert. We simply cannot detect any catalytic advantage for them over the residual activity resulting from determinants unrelated to the triad—see below). Preliminary X-ray analysis of the S221A enzyme indicates no large structural change except for the Ser221 to Ala substitution (R. Bott and M. Ultsch, personal communication). More kinetic and structural data will be necessary however, to substantiate the possible mechanisms discussed above.

The small values of k_{cat} for the active site mutants raise questions regarding protease contaminants or assay artefacts. The following evidence argues strongly against these possibilities. (1) Unlike wild-type subtilisin, the mutant enzymes are not inhibited by phenylmethylsulphonyl fluoride. (2) Although changes in the K_m values are small for these mutants, many are statistically different from wild type (Tables 1, 2). A contamination with helper subtilisin (regardless of amount) would give a constant value for the K_m equal to wild type. (3) Many of the active site mutants differ significantly from each other in k_{cat} and K_m at pH 8.6 (Table 1), which is inconsistent with a constant contaminant. (4) The mutants differ among themselves and wild type in terms of their pH dependence of k_{cat} (Table 2), a result inconsistent with a fixed protease contaminant. (5) Although the kinetic values reported in Tables 1 and 2 are from the same batch of enzyme, most mutant enzymes have been purified more than once. In every repeat case (data not shown) the kinetic values agree within the standard error limits shown ($\leq \pm 15\%$ for k_{cat} and K_m) even though enzyme yields varied, and purification protocols were sometimes slightly modified. (6) The mutants were expressed in an extracellular protease deficient strain of *B. subtilis*, purified on activated thiol sepharose, and judged to be $>99\%$ pure by silver-stained SDS-PAGE. Moreover, further purification of the S24C:H64A enzyme by native gel electrophoresis gave identical kinetic values as the starting material¹².

It is formally possible that the residual activity in some or all of these mutants occurs at a non-specific site(s) distinct from the active site. The following points argue for catalysis at the active site. (1) In some cases the kinetic effects are cumulative for mutagenesis at the active site. For example, the k_{cat} values decrease in the following order: S24C > S24C:D32A > S24C:D32A:H64A > S24C:D32A:H64A:S221A (Table 1). (2) The K_m values are usually not more than twofold above the wild type value suggesting continued strong and specific binding (assuming $K_m \sim K_s$). Furthermore, the active site mutants show a similar pH dependent increase in K_m as wild type subtilisin. (3) The substrate preferences for the S24C:D32A and S24C:S221A enzymes toward two other substrates essentially parallel the wild type enzyme (P. C., unpublished results). The substrate specificity of the S24C:H64A enzyme also parallels the wild type except for a strong preference for His P2 substrates¹² (see below). (4) The activity of the S24C:H64A enzyme is heat denaturable (C. Mitchinson, unpublished results) which indicates that the native protein conformation is critical for catalysis. (5) The residual activity for even the least active mutant is still $>10^3$ fold above the non-enzymatic rate. This catalytic rate is in the range measured for 'good' catalytic antibodies¹⁸⁻²⁰. Taken together these data provide compelling evidence that the residual catalytic activities we have measured are not due to protease contamination, assay artefacts or non-specific catalysis away from the normal active site.

We suggest that the residual activity in the triple mutant is derived from remaining binding determinants which stabilize

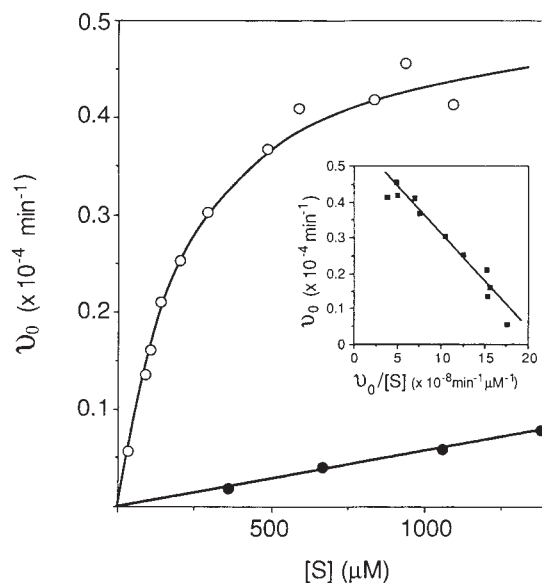


Fig. 2 Initial rate of hydrolysis v_0 ($\Delta A_{410}/\Delta t$) versus the concentration of the substrate *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide [S] in the absence (●) or presence (○) of S24C:D32A:H64A:S221A subtilisin. The background hydrolysis rate (●) was subtracted directly from the rate in the presence of subtilisin to give the enzymatic rate (○). Experiments were performed in 100 mM Tris · HCl, pH 8.60, at $25 \pm 0.2^\circ\text{C}$, as described in Table 1. Insert (■) shows an Eadie-Hofstee plot of the initial rate data.

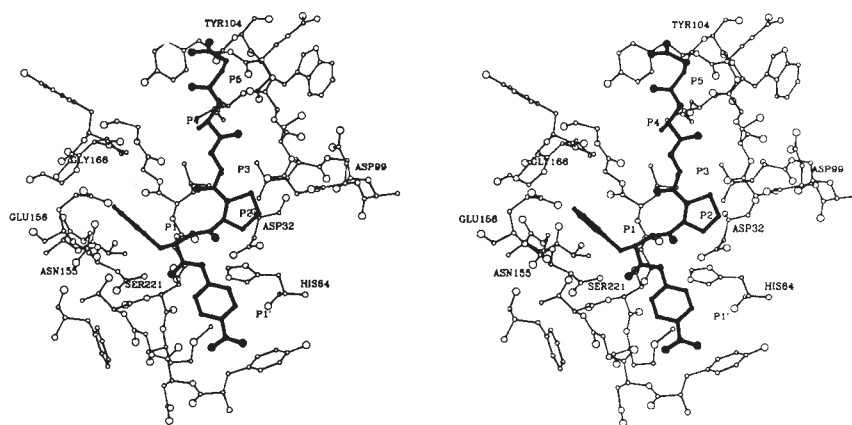
the transition state complex outside the catalytic triad. In fact, previous data show that when the hydrogen bond to Asn155 in the oxyanion binding site (Fig. 1) is disrupted by site-directed mutagenesis, there is a 10^2 to 10^3 drop in k_{cat} with little effect upon K_m ^{14,21}. Additional hydrophobic interactions (Fig. 3) with the P1 substrate side chain⁸ and binding interactions with the P2 to P4 substrate residues^{22,23} are estimated to contribute independently factors of 10 to 100 to k_{cat} . Structural analysis²⁴ suggests there are additional hydrogen bonds in the transition state complex between the NH of Ser221 and the oxyanion, and between the NH of the P1 substrate residue and the carbonyl of Ser125. Deriving the total catalytic contribution from the sum of these individual binding components may lead to overestimation because of their possible interdependence. Nonetheless, our data indicate that some or all of these determinants are important for stabilizing the tetrahedral transition state complex (contributing $>10^3$ to k_{cat}), and are not simply required for positioning the substrate for optimal nucleophilic attack by Ser221.

From an evolutionary point of view, it is extremely unlikely that the catalytic triad arose in one step rather than involving active intermediates. This view is now apparently complicated by the fact that the residues in the catalytic triad function in an extremely synergistic manner. But, assuming that the present-day enzyme is a reasonable model of its ancestor, there are at least two possible mutagenic pathways that give progressive increases in catalytic rate by stepwise introduction of the residues in the triad. In the first pathway, installing Ser221 followed by His64 and then Asp32 gives progressive increases of 8, 9 and 3×10^4 in k_{cat} (Table 1). This progression is even more uniform under alkaline conditions, resulting in increases in k_{cat} of 50, 10 and 5×10^3 (Table 2). A second mutagenic pathway is possible by preferential use of a His P2 substrate (Fig. 3)³⁰ in place of the catalytic His64. We have previously shown that the Ala64 enzyme has a turnover number of $2 \times 10^{-2} \text{ s}^{-1}$ for hydrolysis of a His P2 substrate compared to $8 \times 10^{-6} \text{ s}^{-1}$ for an Ala P2 substrate¹². This catalytic advantage, which we have called 'substrate-assisted catalysis', makes it feasible to reverse the order of introducing His64 and Asp32.

Fig. 3 Stereoview of a model containing the substrate, *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (bold lines and filled atoms), bound to the active site of *B. amyloliquefaciens* subtilisin. Alpha carbons from important enzyme and substrate residues are labelled. In protease substrate nomenclature the substrate may be represented as



where the scissile peptide bond is between the P1 and P1' residues³⁰. The E·S model is based upon a preliminary 2.0 Å X-ray structure of a product bound to subtilisin and the succinyl and *p*-nitroanilide groups were introduced by modelling (R. Bott and M. Ultsch, unpublished data). This model is similar to a previously published complex³¹.



Of course this advantage would apply only to His P2 substrates but would be reasonable if the ancestral enzyme were involved in specific proteolytic processing, for example. Regardless of the exact order of evolutionary events, our mutagenic studies show that inserting catalytic triad residues in a stepwise fashion can produce enzyme intermediates with progressively increased turnover numbers.

In summary, when residues in the catalytic triad are altered separately or together there are large effects on turnover rate, consequent changes in the enzyme mechanism, and only minor effects on the Michaelis constant. The residues in the catalytic triad function in a strongly synergistic fashion and contribute a factor of about 2×10^6 to the total to the catalytic rate enhance-

ment of 10^9 to 10^{10} . The residual activity from complete replacement of the catalytic triad is not a contaminant or other artefact, but results from transition state stabilization from contacts outside the catalytic triad. Finally, despite the synergy between the catalytic triad residues, their sequential introduction is reasonable in terms of both evolution and function.

We thank Dr Rick Bott for help in preparing Fig. 3 and sharing unpublished X-ray coordinates, Dr Polly Moore and Ann Benninger for assistance in data handling, the organic chemistry group at Genentech for synthesis of oligonucleotides, and Drs Tony Kossiakoff, Jack Kirsch and Ron Wetzel for helpful comments on this manuscript.

Received 19 January 1988; accepted 22 February 1988.

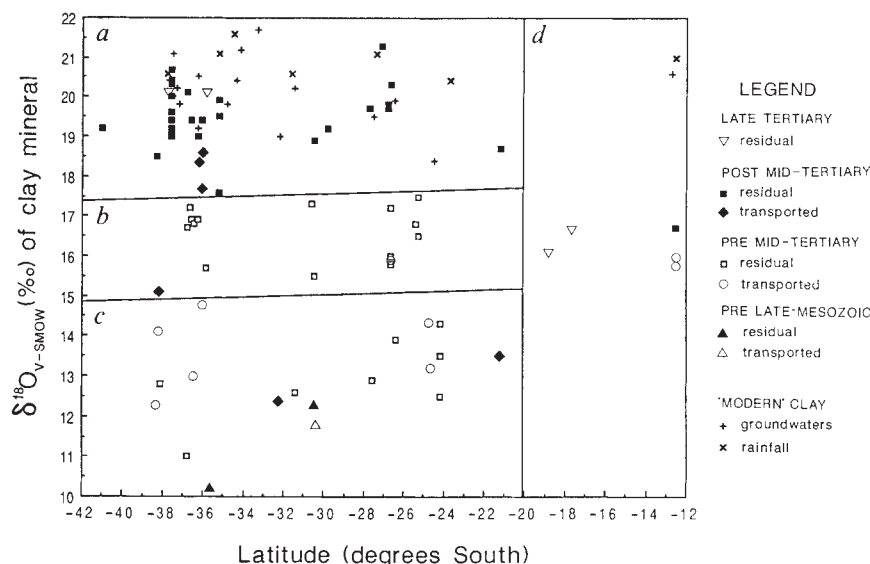
1. Stroud, R. M. *Scient. Am.* **131**, 74-88 (1974).
2. Kraut, J. A. *Rev. Biochem.* **46**, 331-358 (1977).
3. Fink, A. L. in *Enzyme Mechanisms* (eds Page, M. I. & Williams, A.) 159-177 (Roy. Soc. Chem. 1987).
4. Kossiakoff, A. A. in *Biological Macromolecules and Assemblies* Vol. 3 (eds Jurnak, F. A. & McPherson, A.) 370-412 (1987).
5. Zoller, M. J. & Smith, M. *Nucleic Acids Res.* **10**, 6487-6500 (1982).
6. Carter, P., Bedouelle, H. & Winter, G. *Nucleic Acids Res.* **13**, 4431-4443 (1986).
7. Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A. & Chen, E. Y. *Nucleic Acids Res.* **11**, 7911-7925 (1983).
8. Estell, D. A. *et al. Science* **233**, 659-663 (1986).
9. Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. & Estell, D. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1219-1223 (1987).
10. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5167-5171 (1987).
11. Power, S. D., Adams, R. M. & Wells, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3096-3100 (1986).
12. Carter, P. & Wells, J. A. *Science* **237**, 394-399 (1987).
13. Wells, J. A. & Powers, D. B. *J. biol. Chem.* **261**, 6564-6570 (1986).
14. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. *Phil. Trans. R. Soc. A* **317**, 415-423 (1986).

15. Gutfreund, H. & Sturtevant, J. M. *Biochem. J.* **63**, 656-661 (1956).
16. Craik, C. S., Rocznik, S., Largman, C. & Rutter, W. J. *Science* **237**, 909-913 (1987).
17. Sprang, S. *et al. Science* **237**, 905-909 (1987).
18. Tramontano, A., Janda, K. D. & Lerner, R. A. *Science* **234**, 1566-1570 (1986).
19. Pollack, S. J., Jacobs, J. W. & Schultz, P. G. *Science* **234**, 1570-1573 (1986).
20. Napper, A. D., Benkovic, S. J., Tramontano, A. & Lerner, R. A. *Science* **237**, 1041-1043 (1987).
21. Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H.-Y. & Poulos, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3743-3745 (1986).
22. Morihara, K., Oka, T. & Tsuzuki, H. *Arch. Biochem. Biophys.* **138**, 515-525 (1970).
23. Morihara, K., Oka, T. & Tsuzuki, H. *Biochem. biophys. Res. Commun.* **35**, 210-214 (1969).
24. Robertus, J. D., Kraut, J., Alden, R. A. & Birktoft, J. J. *Biochemistry* **11**, 4293-4303 (1972).
25. Wells, J. A., Vasser, M. & Powers, D. B. *Gene* **34**, 315-323 (1985).
26. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
27. Yang, M. Y., Ferrari, E. & Henner, D. J. *J. Bact.* **160**, 15-21 (1984).
28. DelMar, E. G., Langman, C., Brodrick, J. W. & Goekas, M. C. *Analyt. Biochem.* **99**, 316-320 (1979).
29. Matsubara, H., Kasper, C. B., Brown, D. M. & Smith, E. L. *J. biol. Chem.* **240**, 1125-1130 (1965).
30. Schechter, I. & Berger, A. *Biochem. biophys. Res. Commun.* **27**, 157-162 (1967).
31. Robertus, J. D. *et al. Biochemistry* **11**, 2439-2449 (1972).

Erratum Oxygen isotope dating of the Australian regolith

Michael I. Bird & Allan R. Chivas
Nature **331**, 513-516 (1988)

IN this letter, Fig. 1 as printed is too small to allow the symbols to be properly differentiated. The figure is reprinted here with enlarged symbols. In addition, line 13 in the left-hand column on page 515 has become garbled by an error in a line correction. The first sentence in that paragraph should read: "Isotopic results obtained from residual clays (collected *in situ* from regolith profiles) of post-mid Tertiary age have $\delta^{18}\text{O}$ values between 17.5 and 21.3‰, with the exception of samples from field *d* (representing latitudes north of $\sim 20^\circ\text{S}$) which have anomalously low values." □

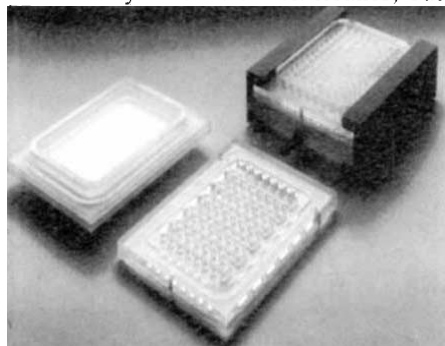


New product arrivals for spring

The newest laboratory products for spring include reagents for making probes for cellular oncogenes, a chamber for oxygen consumption experiments and a device for cutting Parafilm.

THE BladeRunner from Kontes makes **Parafilm cutting** easier (*Reader Service No. 100*). With the BladeRunner, Parafilm can be cut in squares or strips, with one hand. The BladeRunner's floating blade prevents the cut edge of the Parafilm from sticking to the backing, so uniform, straight cuts can be made at any length. The cutting head automatically returns after a cut for convenient repetitive dispensing, and the blades can be replaced with standard single-edge razor blades. The BladeRunner accommodates both two- and four-inch rolls of Parafilm, and costs \$165 (US).

Millipore has announced the availability of its MilliBlot systems for **nucleic acid and protein transfers** (*Reader Service No. 101*). The MilliBlot systems are designed for use with Millipore's Immobilon family of transfer membranes, but



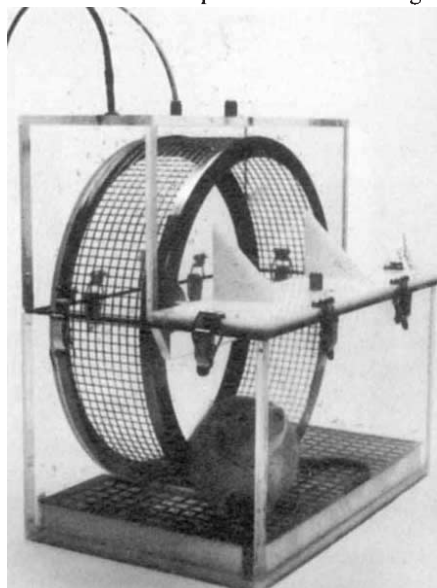
The modular MilliBlot from Millipore.

may also be used with most other commonly used transfer membranes. The \$900 (US) MilliBlot systems accommodate gels and membranes of up to 10.5 cm × 7 cm. The modular systems allow nucleic acid transfer, 96-well dot blotting, and 48-well slot blotting with a simple change of module. Millipore says transfers can be completed in as little as 30 minutes, and no power source or lengthy overnight transfers are required.

Sigma has a specialized catalogue on **cell culture reagents** that's just off the presses, to accompany its comprehensive annual catalogue (*Reader Service No. 102*). The catalogue lists products ranging from powdered and liquid media, to more unique products such as Sigma's hybridoma-tested reagents and cell culture-tested biochemicals. An extensive group of serum replacements, media for serum-free applications and numerous varieties of sera are described, along with items such as microcarrier beads, pipettes, soaps and detergents. The catalogue also

includes a wealth of technical information and application methods.

For prolonged **oxygen consumption experiments**, Columbus Instruments International Corp. has a new airtight



An exercise wheel is included in the chamber. acrylic chamber for small laboratory animals (*Reader Service No. 103*). The chamber is equipped with a voluntary exercise wheel and its size conforms to US National Institutes of Health welfare guidelines for prolonged animal housing. The exercise wheel is made of stainless steel, and has an electronic counter for keeping track of the number of rotations. One to four chambers may be connected to Columbus Instruments' Oxymax oxygen consumption computer for the collection of data on oxygen consumption, CO₂ production, respiratory exchange ratio and animal ambulatory and exercise activity. The chamber's dimensions are 15.5 × 9.75 × 15.75 inches, and each costs \$1,990 (US).

Plasmid DNAs for the preparation of **RNA probes for cellular oncogenes** and prostaglandin synthase are now available from Oxford Biomedical Research (*Reader Service No. 104*). The probes can be used for the identification and localization of human *c-myc*, *N-myc* and *v-myb* oncogenes and their transcription products, and for prostaglandin synthase. The plasmid DNAs direct the synthesis of RNA+ or RNA- probes using bacteriophage SP6 and T7 RNA polymerases, respectively. Biotinylated *c-myc*, *N-myc* and *v-myb* RNA+ and RNA- probes have

also been developed. Oxford Biomedical sells the plasmid DNAs for \$95 for 5 µg and \$180 (US) for 10 µg. The company plans to offer several more plasmid DNAs for cellular oncogenes in the near future.

Setting it straight

The photograph published with a review of the synthetic peptide service offered by Applied Biosystems (*Nature* **331**, 462; 1988) was printed in error. The photo depicts the OPC cartridges sold by the company. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE - HOMOGENEOUS -
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIABORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.49

SINGER
INSTRUMENTS

**Cell
Injection?**
Singer MK III
Singer Microforge
Singer Instrument Co. Ltd.

Treborough Lodge, Roadwater, Watchet,
Somerset, England, TA23 0QL.
Tel: Washford (0984) 40226
Telex: 337492 Comcab G

Reader Service No.50

Low-tech US jobs boom?

Richard Pearson

Twenty-one million new jobs are expected in the United States by the year 2000, and unemployment is likely to fall to 6 per cent.

THE Bureau of Statistics is projecting 21 million new jobs in the United States before the year 2000, an increase of nearly 20 per cent. The service sector will expand, but agriculture will continue its long-term employment decline; the growth in manufacturing output is expected to be overtaken by productivity gains leading to further employment decline, a reversal of the recent improved employment prospects. Among the goods-producing industries, only construction is showing long-term job growth, with an increase expected of 900,000 jobs almost matching the expected decline of 800,000 jobs in manufacturing. The proportion of jobs in the service sector is expected to rise from 3 in 4 in 1986 to almost 4 out of 5 in the year 2000. When account is taken of the 18 per cent growth in the labour force that the bureau is projecting, the unemployment rate will fall to 6 per cent. This assumes economic growth of about 2.4 per cent a year, slightly less than the 2.5 per cent achieved over the past 14 years; it assumes also that exports will increase their share of GNP, imports will retain a static share and defence and state and local government will have a reduced share although government will

Sectoral change 1986–2000 mid-point estimates (thousands)

Mining	–59
Construction	+890
Manufacturing	–835
Transport & utilities	+475
Wholesale Trade	+1,531
Retail Trade	+4,857
Finance etc	+1,620
Services	+10,014
Government	+1,618
Agriculture	–335
Self employment	+1,630

increase its rate of growth. Alternative scenarios are also presented but the focus here is on the mid-point trends and those of importance to the scientifically and technologically qualified.

Although the manufacturing sector is the traditional employer of scientific and technological manpower, its contraction is not a picture of complete gloom. For example, employment in computer and office equipment manufacture is expected to grow by about 85,000 jobs, or 20 per cent, although this represents a considerable slowdown in historic terms. Another high-technology area is electronics, where

employment is expected to remain static, although with a considerable expansion in companies producing health-related products and those producing satellites, fibre optics, and laser and broadcast equipment for private users, which will counteract the decline in defence demand. The other growth sector in manufacturing is pharmaceuticals, where advances in biomedical research are leading to new drugs and diagnostic tools; the growth in the elderly population, who are high spenders on these products, will further boost demand, employment growing by nearly 8 per cent to reach 224,000 by the year 2000.

It is the service sector, however, that will be the job generator of the future, matching broader trends in the developed world (*Nature* 331, 464; 1988, and 332, 96; 1988). The most rapidly growing service sector will be computer and data-processing services: growth in demand for specialized systems by government and business will continue to boost the demand for systems design and analysis, programming and software development skills. Employment in this sector is expected to double to 1.2 million by the end of the century. With the ageing population boosting demand for health care, employment in health services is expected to expand by over 3 million, a nearly 50 per cent increase. But cost-containment policies are halting, at least temporarily, the growth in demand for hospital facilities and promoting outpatient services, so employment growth is expected to centre on the offices of health practitioners. Other rapidly growing sectors will be business and personal services, including the temporary help phenomenon, and the retail and wholesale trades.

How then do these sectoral trends affect the job prospects for scientists and technologists? For life scientists, job growth will be in line with the national average at 21 per cent, or 30,000 jobs, building on the advances in genetic research in areas such as new medicines, plant and animal developments, and diagnostic techniques for genetic defects. Employment of physical scientists is expected to expand rather more slowly at 13 per cent, with 24,000 jobs being added because of military and private research and development, and with new prospects in advanced areas such as laser research and high-energy physics. Jobs for computer systems

analysts will show far higher rates of growth, with 250,000 more jobs, an increase of 76 per cent. The expansion in the computer industry will account for about half this growth.

In engineering, electrical/electronic engineers lead the jobs boom with a net gain of 192,000 jobs, a 48 per cent growth, building on the expansion in communications, computing and electronics industries as they seek to remain competitive by using these skills to design new products and to improve existing products and production processes. A further 76,000 jobs are expected for mechanical engineers,

High-tech job growth in selected occupations 1986–2000 (thousands)

Computer science	251
Computer programming	335
Electrical engineering	192
Physicians/surgeons	188
Mechanical engineering	76
Life sciences	30
Physical sciences	24

Source: US Bureau of Labor Statistics

not only in manufacturing but also in service companies and the temporary-help sector. The expansion in health care will add 188,000 jobs for physicians and surgeons, up by 38 per cent. Technicians will also gain in all these sectors, with the main increase being for health technicians (663,000 jobs), computer programmers (335,000 jobs) and in engineering and science (285,000 jobs). But college and university are expected to fall by 4 per cent because of a decline in enrolments.

In themselves, these figures give support to a thesis of a high-technology future where job opportunities will go to the highly qualified. But in reality, the real growth in job opportunities will be in low-technology areas such as retail sales where jobs will grow by over one million; there will be 750,000 more jobs for waiters and 600,000 for cleaners and janitors, but few of these jobs will offer chances of promotion and advancement. There will be an increase in the proportion of jobs normally filled by college leavers, but they will still only account for about 1 in 4 jobs at the end of the century and low technology will remain predominant. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.